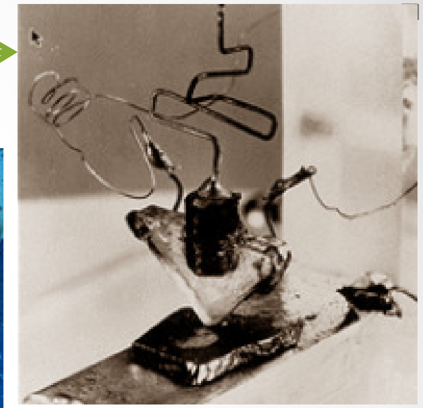
molecular programming

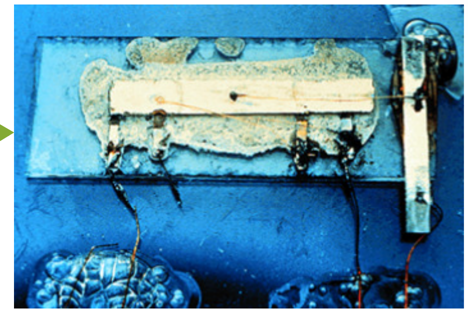
Luca Cardelli
Microsoft Research
Lab Tutorial
2010-02-11

Smaller and Smaller

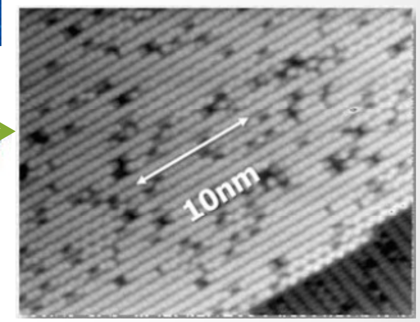
Dec. 23, 1947. John Bardeen and Walter Brattain show the first working transistor.



September 1958. Jack Kilby builds the first integrated circuit.

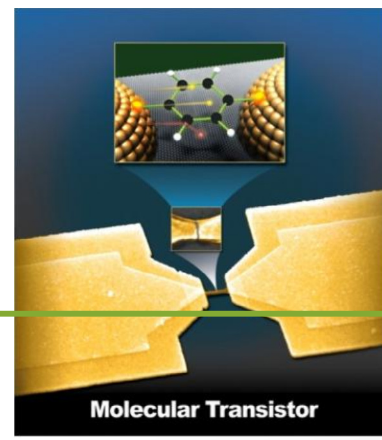


Jan 30, 2010. Intel and Micron announce 25nm NAND flash.



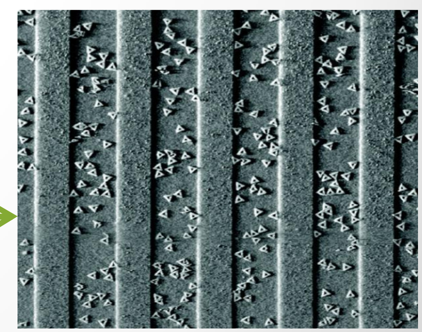
Scanning tunneling microscope image of a silicon surface showing 10nm is ~20 atoms across

Dec. 24, 2009. Working transistor made of a single molecule.



Observation of molecular orbital gating. *Nature*, 2009; 462 (7276): 1039

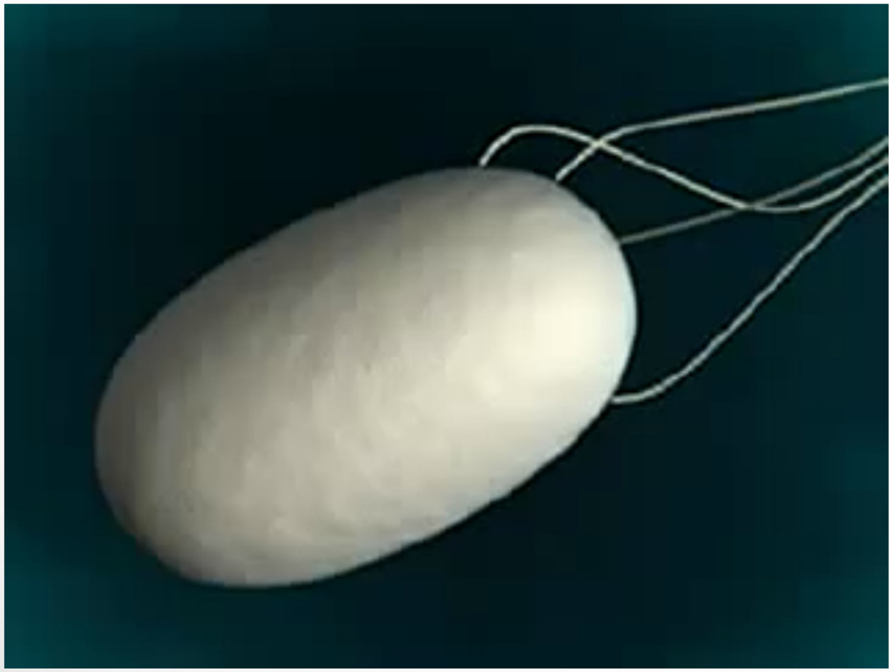
The race is on for *molecular scale integrated circuits*.



Placement and orientation of individual DNA shapes on lithographically patterned surfaces. Nature Nanotechnology 4, 557 – 561 (2009).

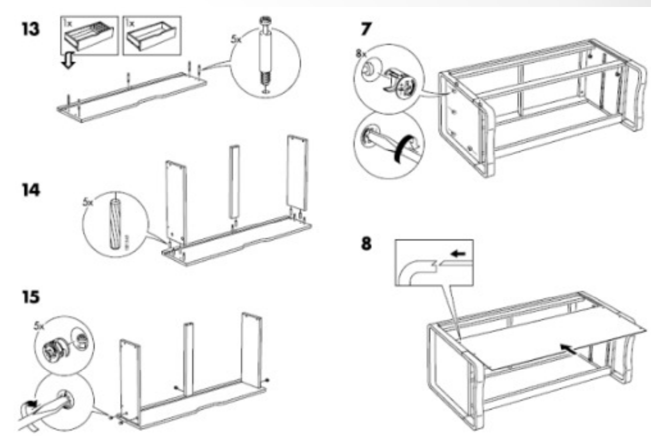
Building The Small*est* Things

- How do we build structures that are by definition smaller than your tools?
- Basic answer: you can't. Structures (and tools) should build themselves!
- By *programmed self-assembly*.

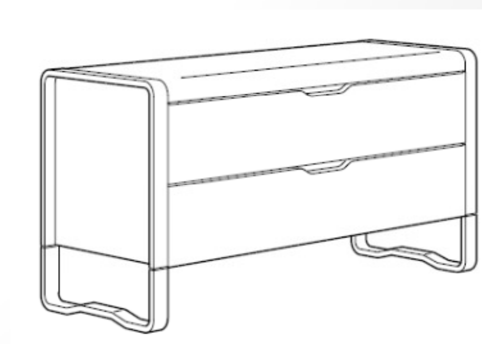


Molecular IKEA

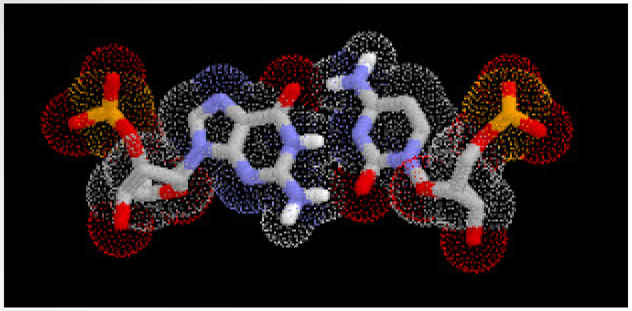
- Nature can self-assemble.
  Can we?
- *"Dear IKEA, please send me a chest of drawers that assembles itself."*
- We need a magical material where the pieces are pre-programmed to fit into to each other.
- At the molecular scale many such materials exist; let's pick one…



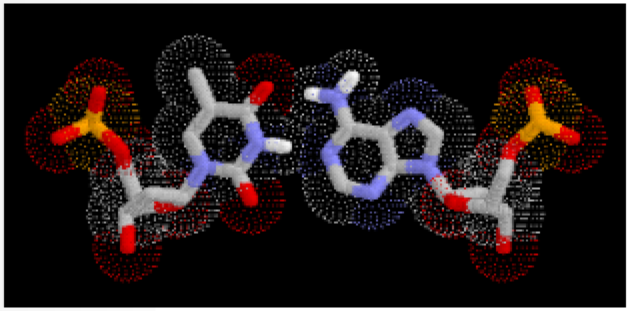
Add water



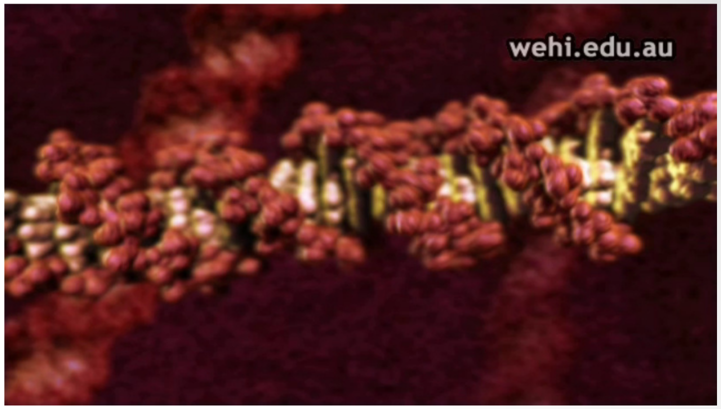
DNA

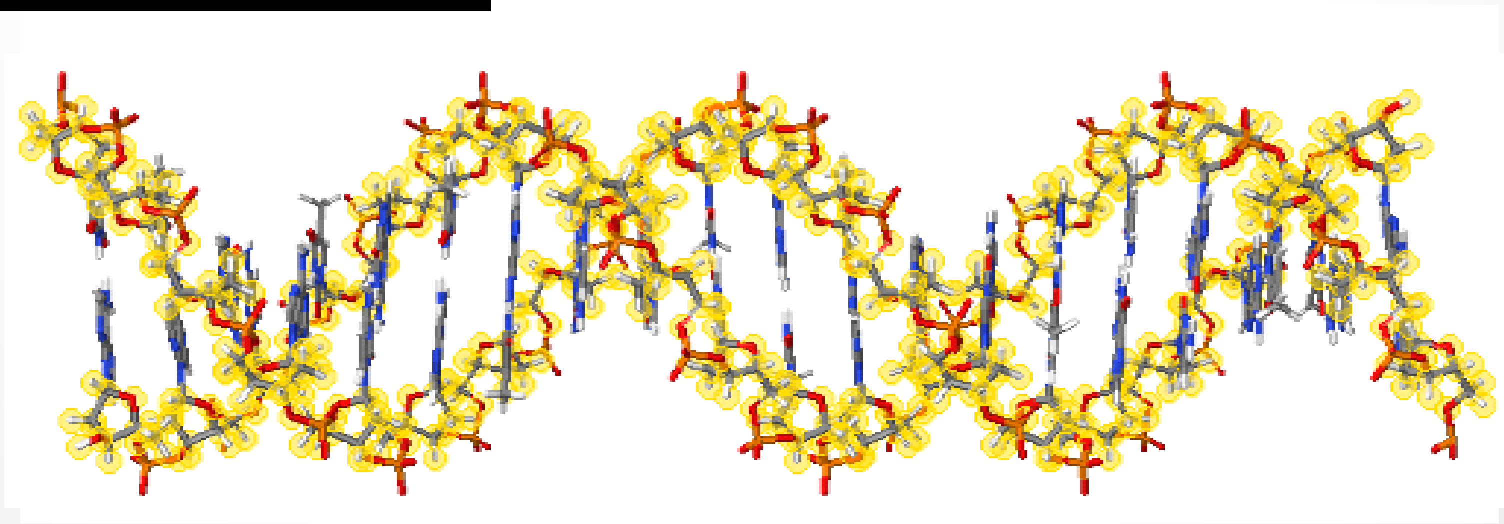


GC Base Pair
Guanine-Cytosine



TA Base Pair
Thymine-Adenine





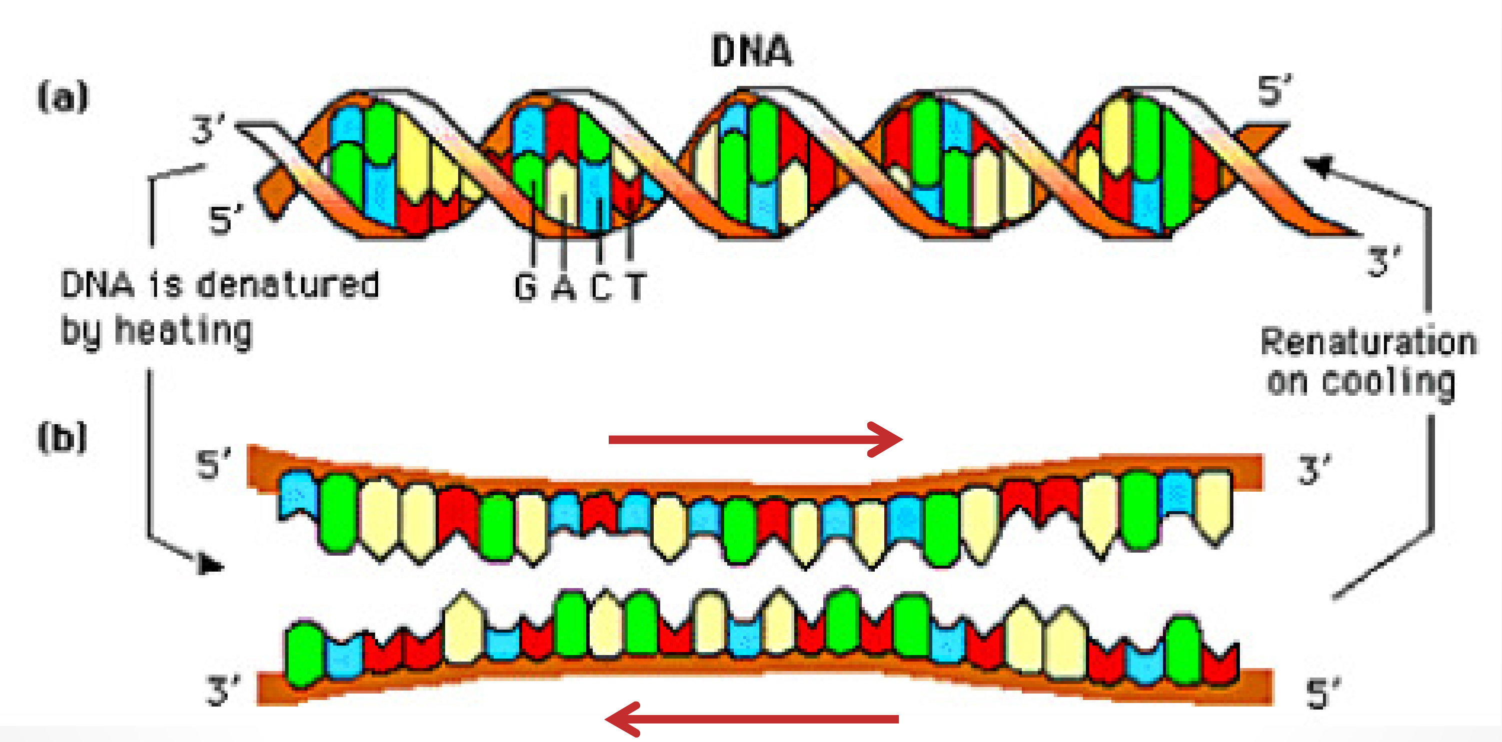
Sequence of Base Pairs (GACT alphabet)

[Interactive DNA Tutorial](http://www.biosciences.bham.ac.uk/labs/minchin/tutorials/dna.html)

(http://www.biosciences.bham.ac.uk/labs/minchin/tutorials/dna.html)

ssDNA

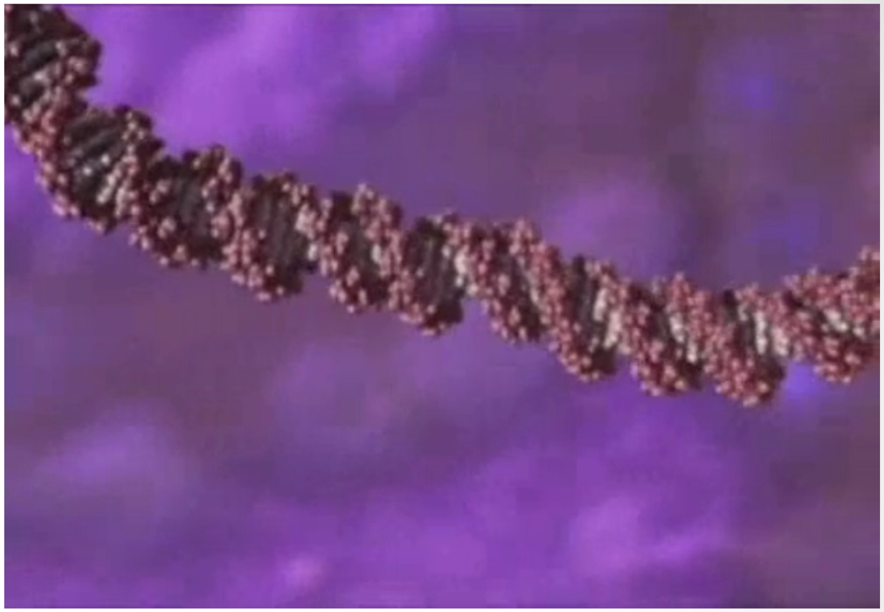
Double-stranded DNA



Single-stranded DNA has an orientation
Each strand spells a GACT sequence
The two strands have *opposite* orientations

Robust, and *Long*

- DNA in each human cell:
  - 3 billion base pairs
  - 2 meters long, 2nm thick
  - folded into a 6μm ball
  - 750 MegaBytes
- A huge amount for a cell
  - Every time a cell replicates it has to copy *2 meters of DNA* reliably.
  - To get a feeling for the scale disparity, compute:
- DNA in human body
  - 10 trillion cells
  - 133 Astronomical Units long
  - 7.5 OctaBytes
- DNA in human population
  - 20 million light years long

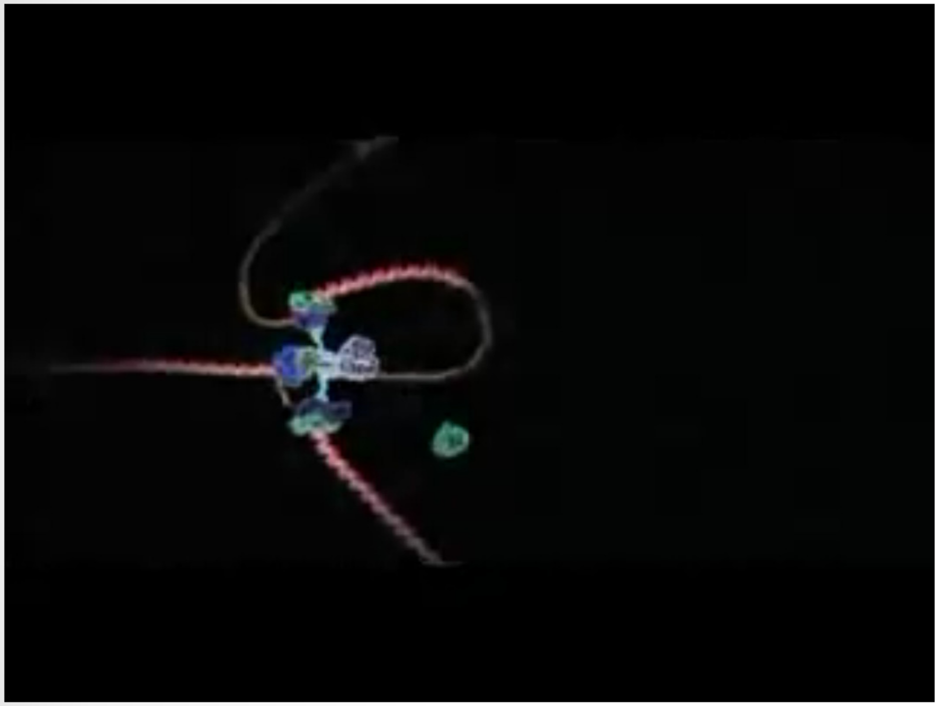


DNA wrapping into chromosomes

Andromeda Galaxy
2.5 million light years

Zipping Along

- DNA can support structural and computational complexity.



DNA replication in *real time*

In Humans: 50 nucleotides/second
Whole genome in a few hours (with parallel processing)

In Bacteria: 1000 nucleotides/second
(higher error rate)

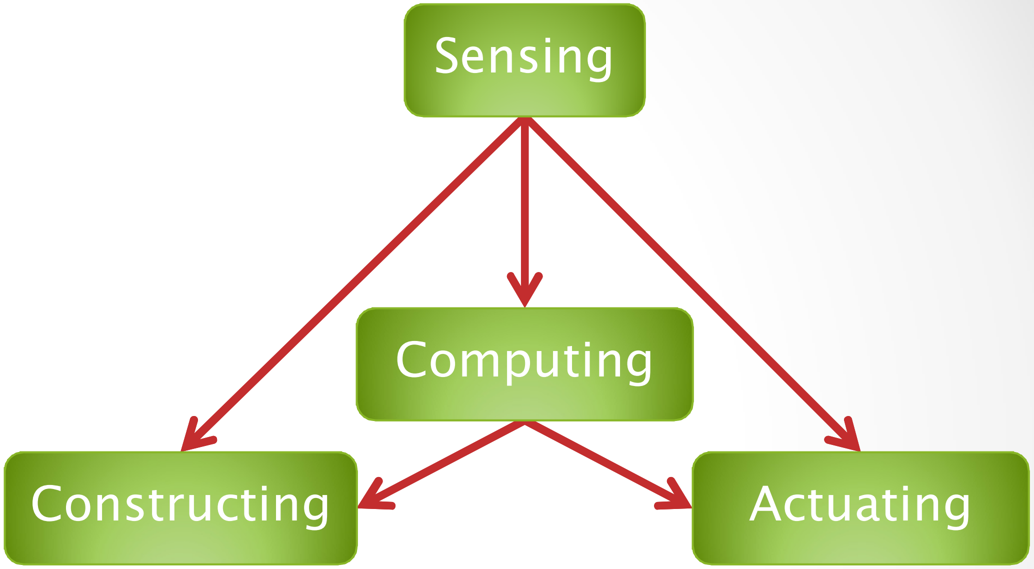
DNA transcription in *real time*

RNA polymerase II: 15-30 base/second

Drew Berry
http://www.wehi.edu.au/wehi-tv

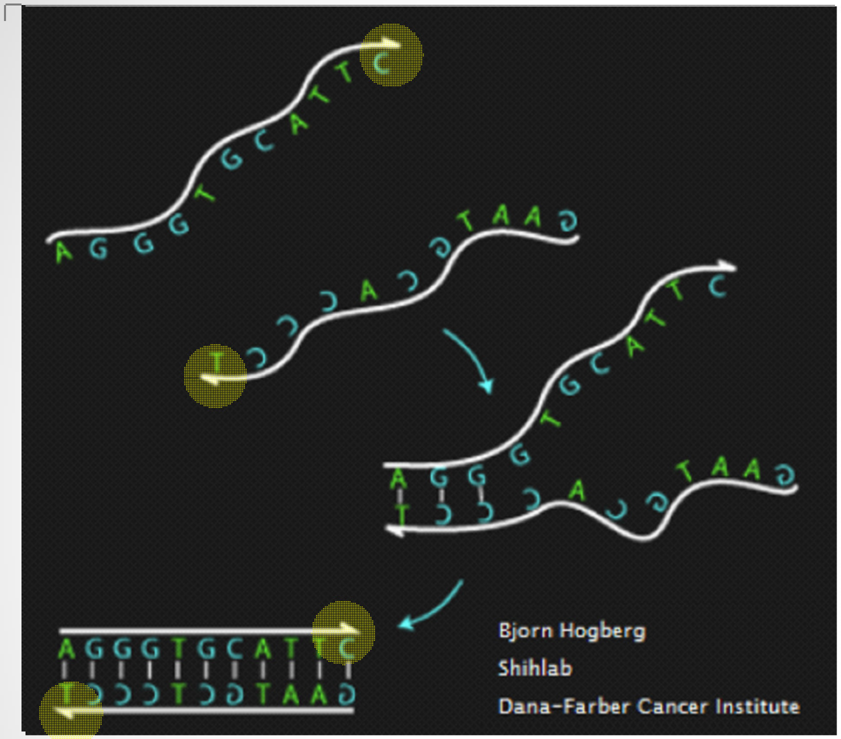
Nanoscale Engineering

- Sensing
  - Reacting to forces
  - Binding to molecules
- Actuating
  - Releasing molecules
  - Producing forces
- Constructing
  - Chassis
  - Growth
- Computing
  - Signal Processing
  - Decision Making

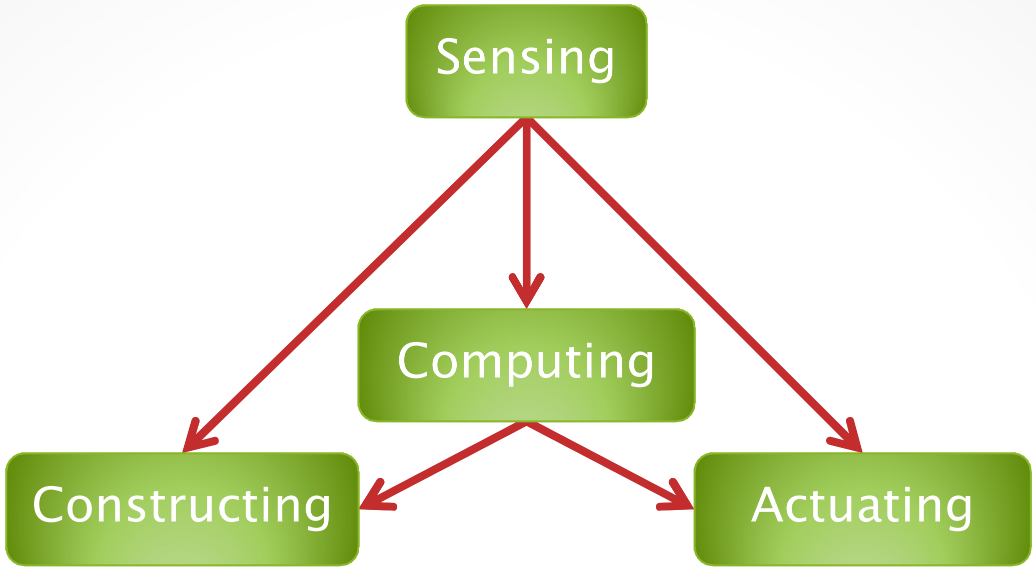


Nucleic Acids (DNA/RNA) can do all this, and interface to biological structures.

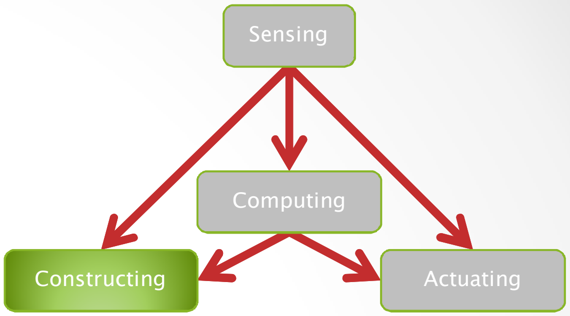
Hybridization



- Strands with **opposite orientation** and **complementary base pairs** stick to each other (Watson-Crick duality).
- This is all we are going to use
  - We are not going to exploit DNA replication, transcription, translation, **restriction and ligation enzymes**, etc., which enable other classes of tricks.



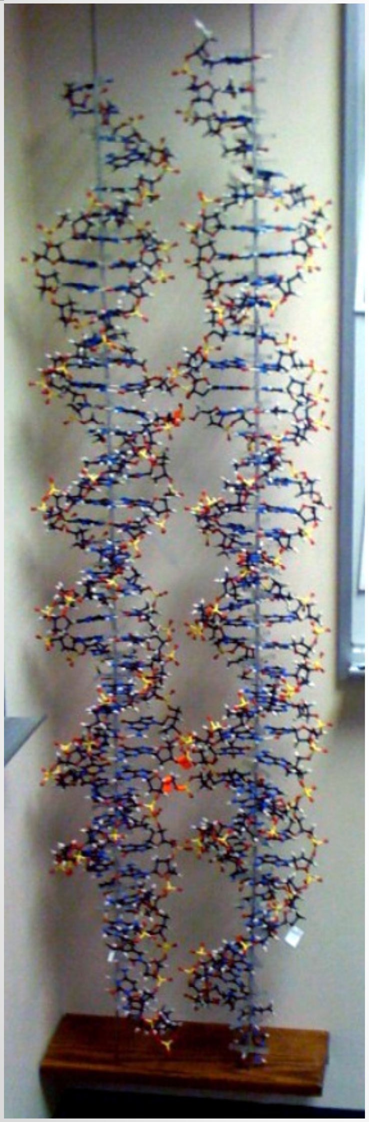
Hybridization Tricks

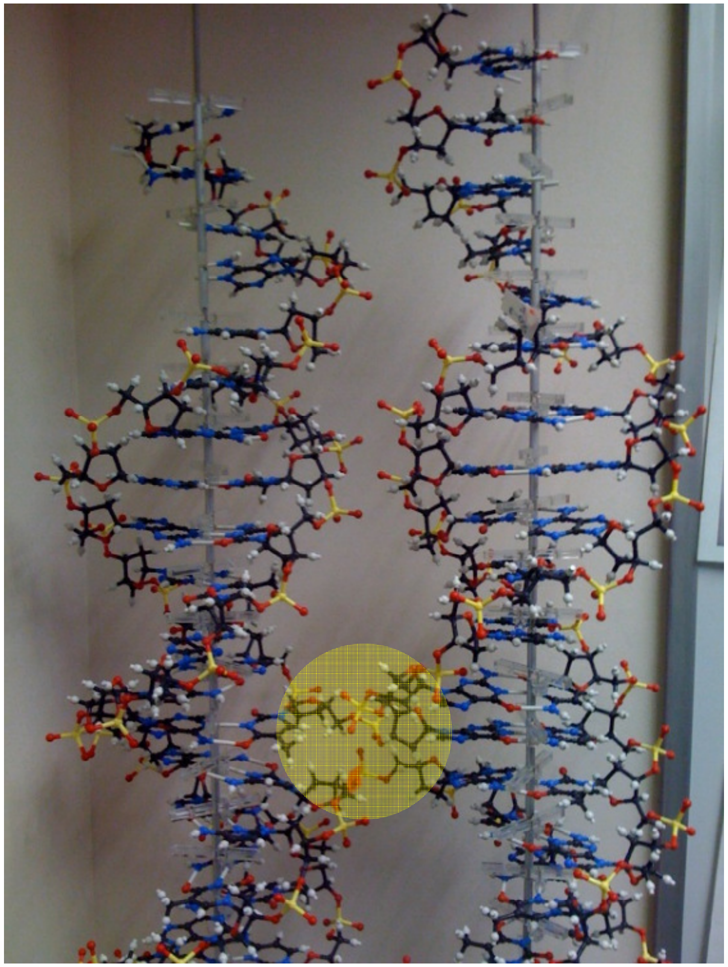


Constructing

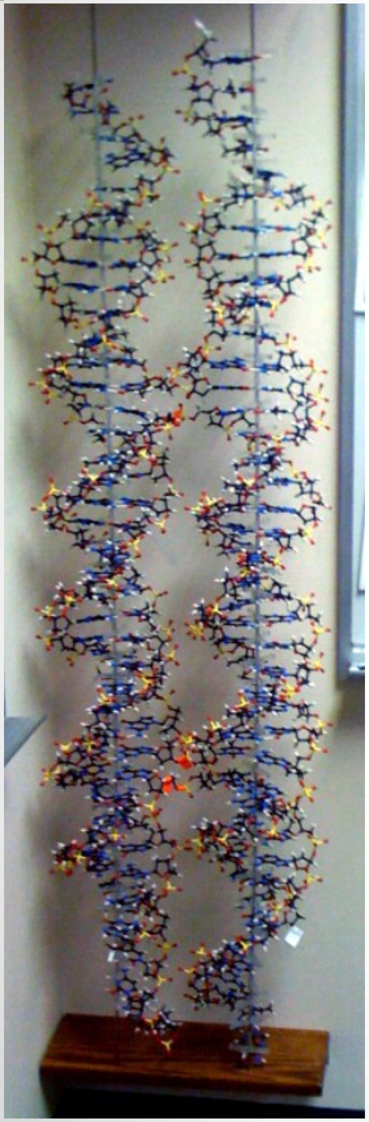
• • •

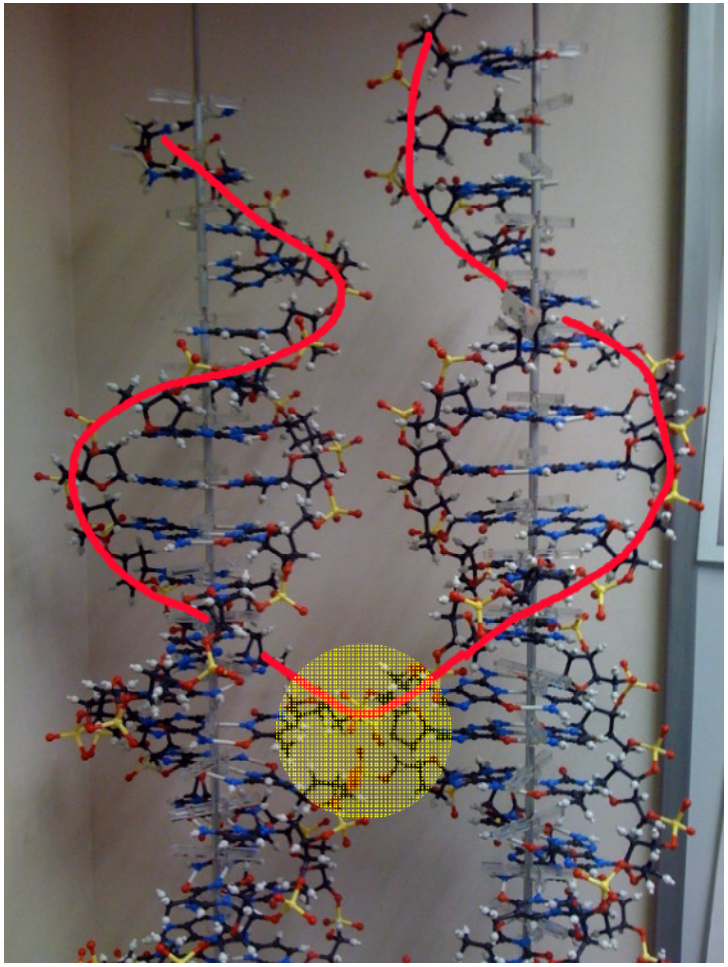
Crosslinking



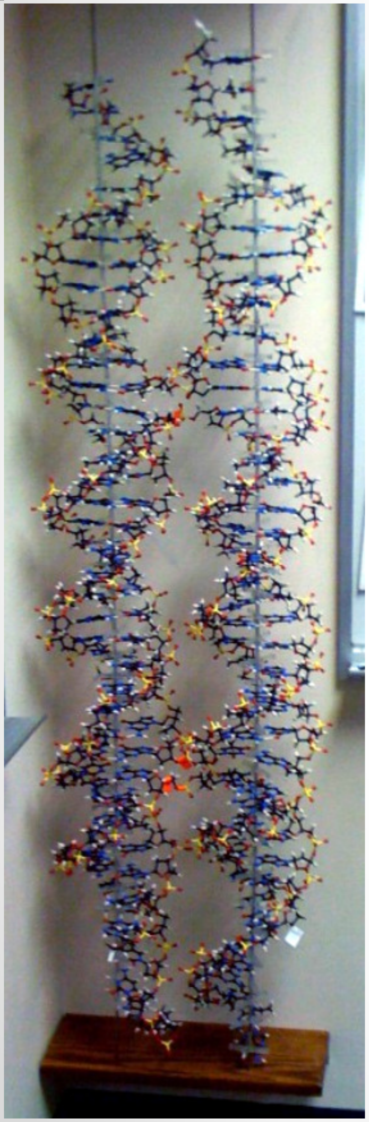


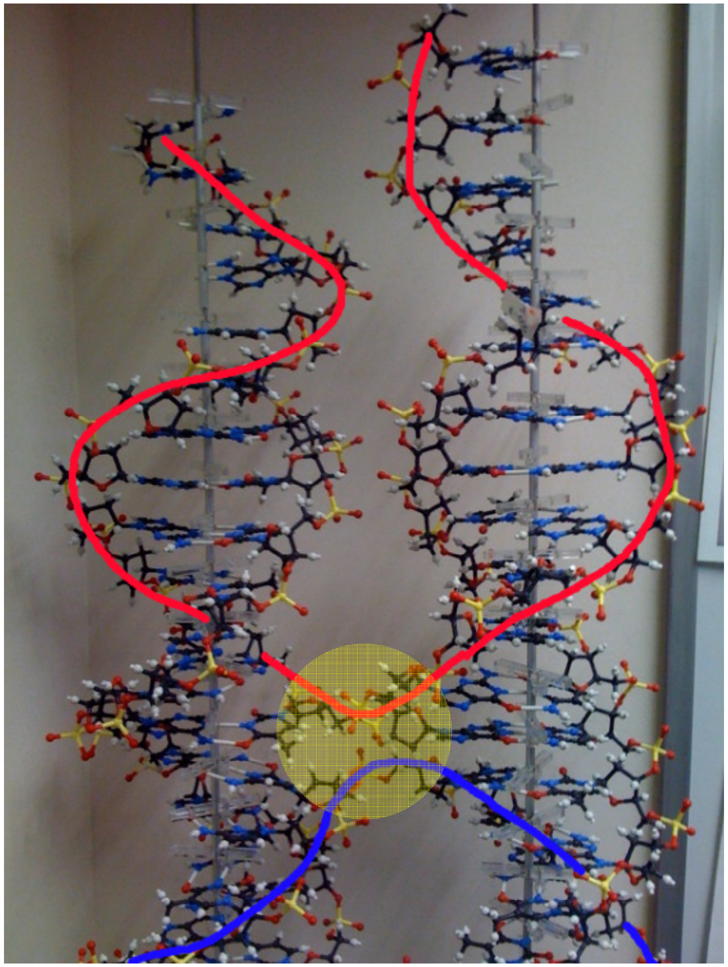
Crosslinking



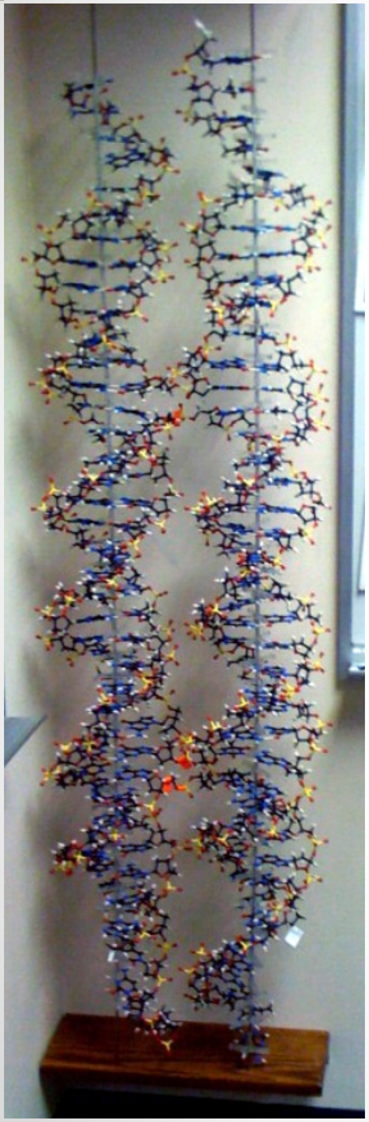


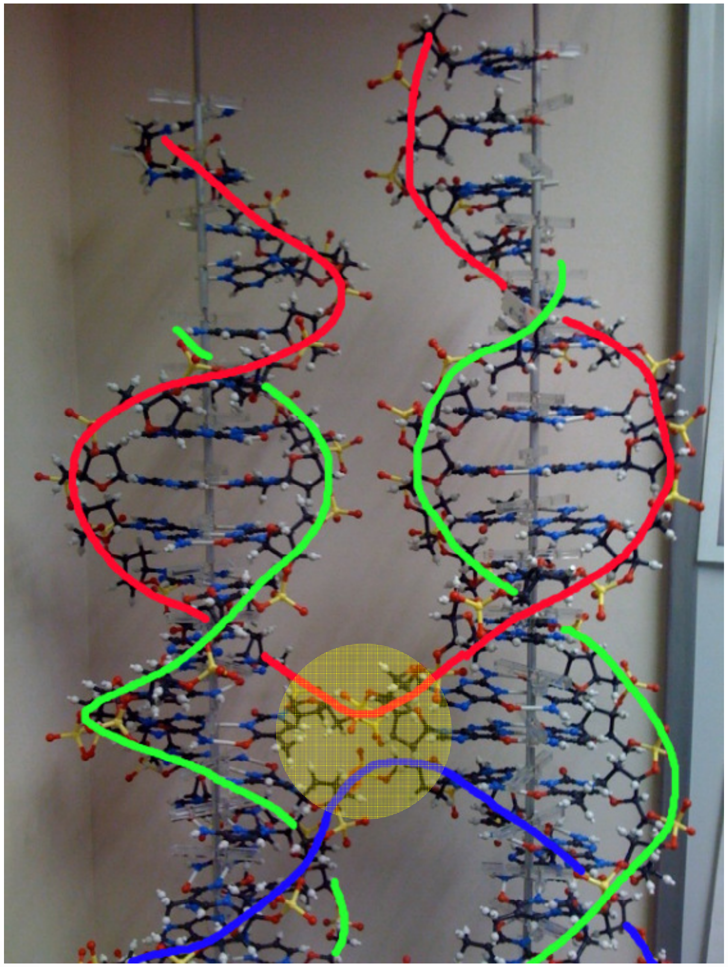
Crosslinking



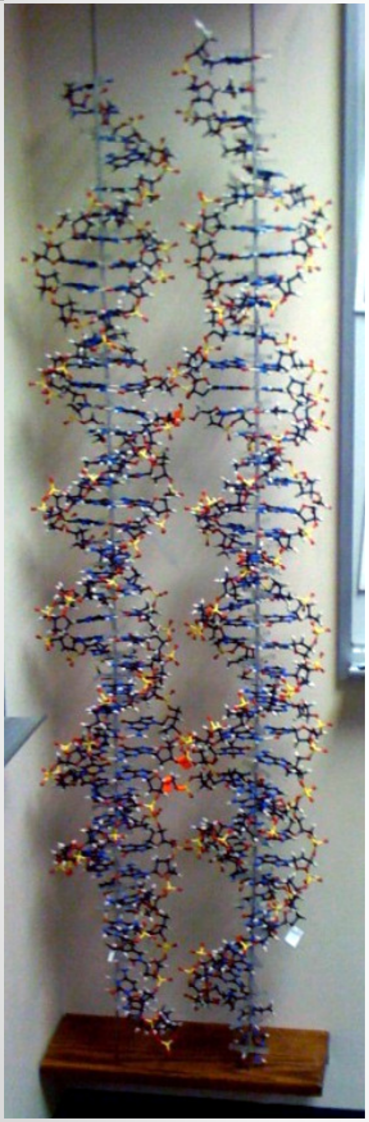


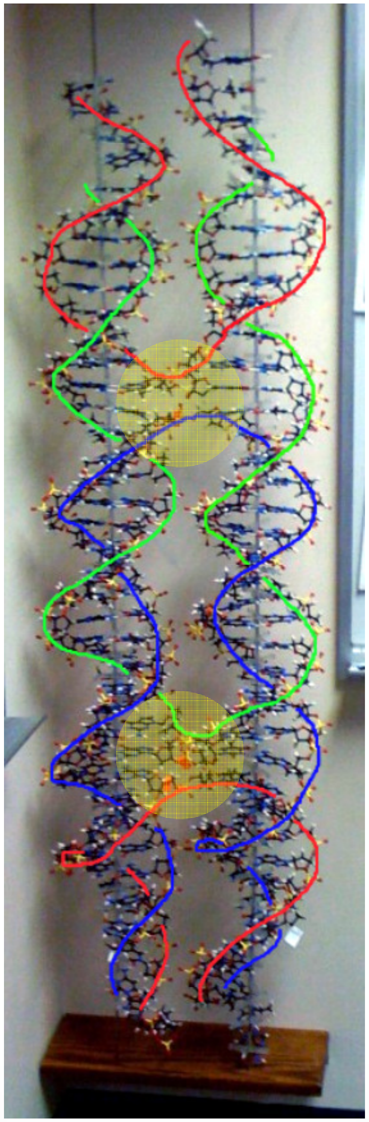
Crosslinking



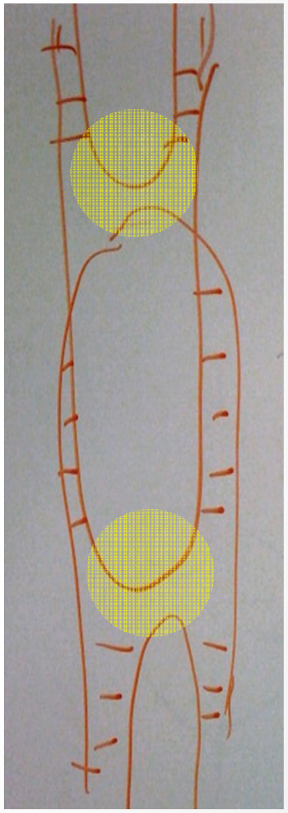


Crosslinking



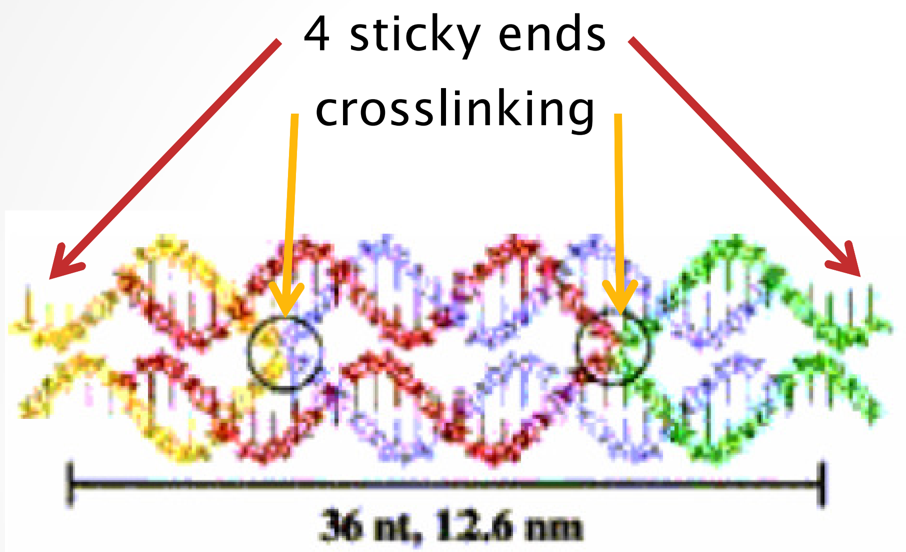


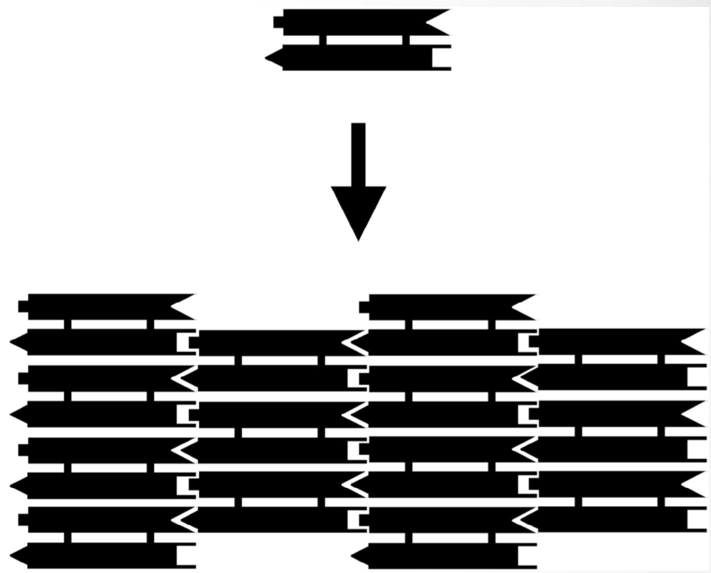
In nature, crosslinking is deadly (blocks DNA replication).

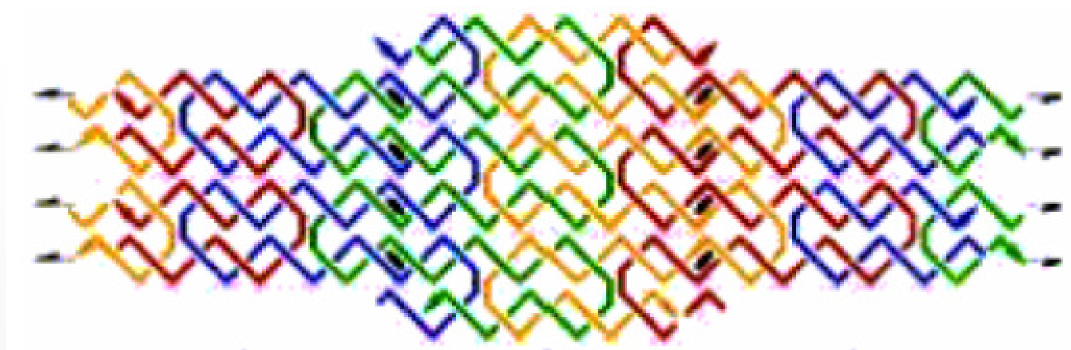


In engineering, crosslinking is the key to using DNA as a construction material.

DNA Tiling



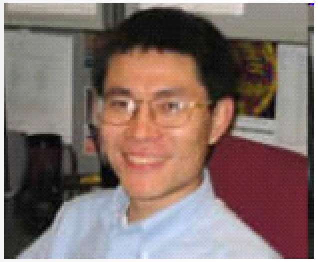




Construction and manipulation of DNA tiles in free space

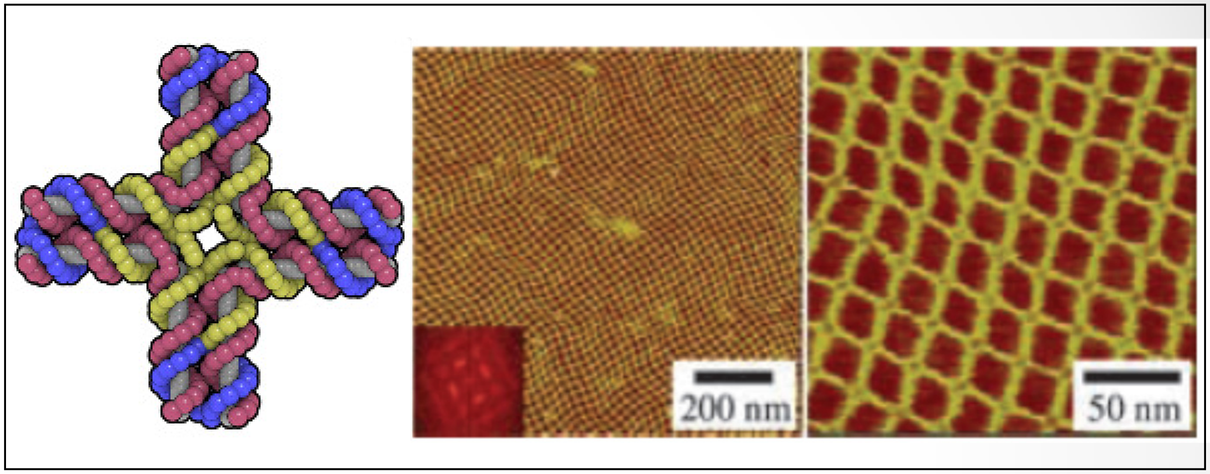
Pankhudi

2D DNA Lattices

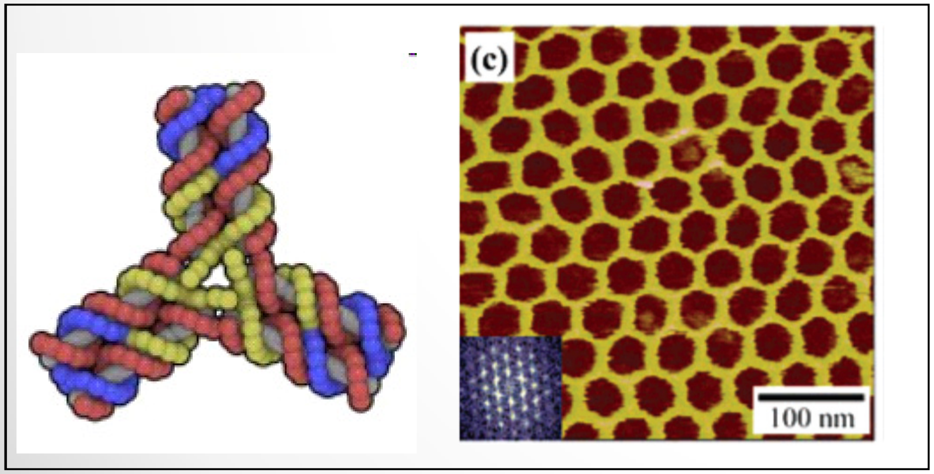


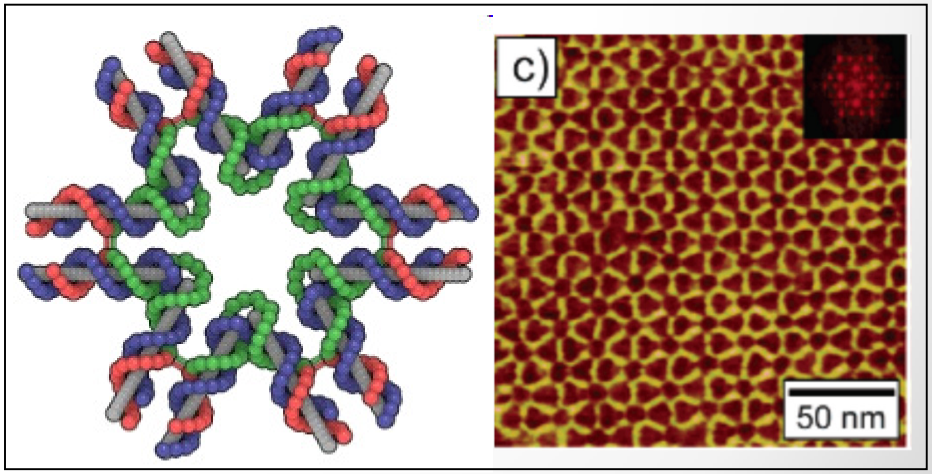
Chengde Mao

Purdue University, USA



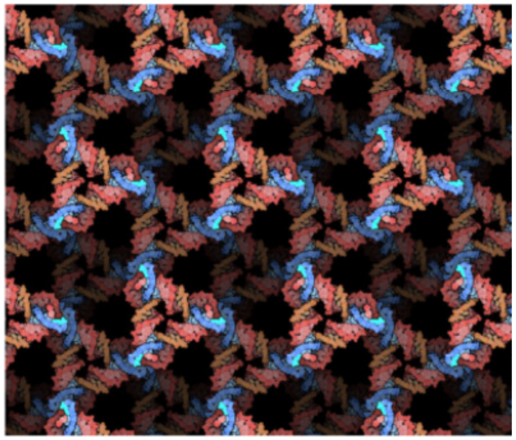
N-point Stars





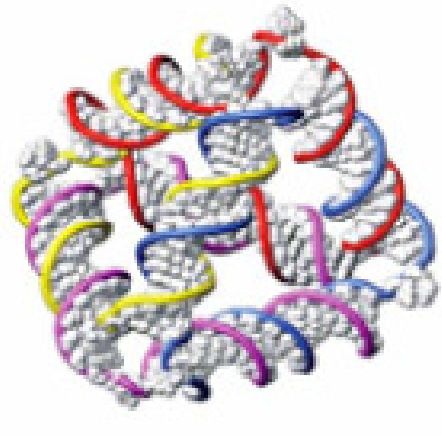
3D DNA Structures

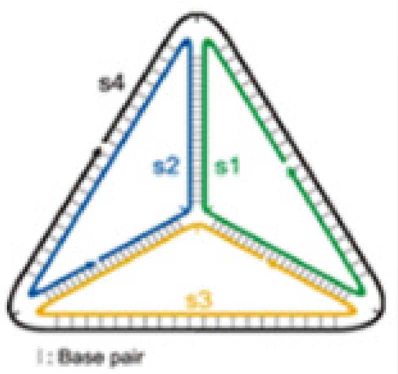
Ned Seeman
NYU



3D Cyrstal

AndrewTuberfield
Oxford





Tetrahedron

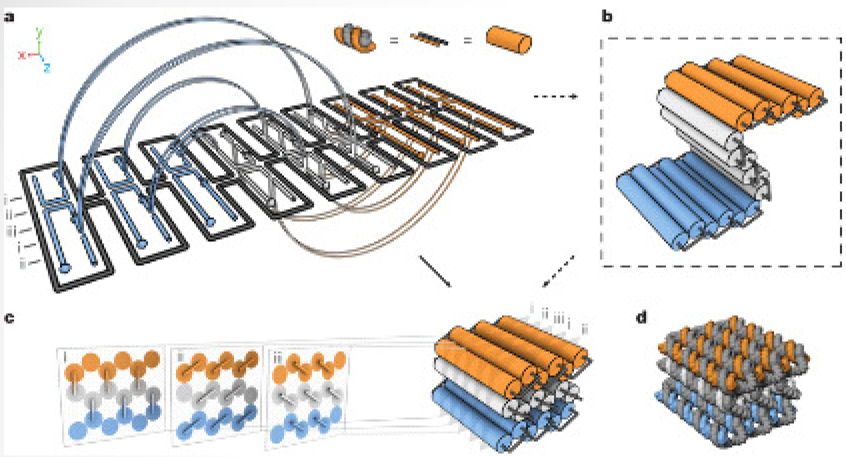
CADnano

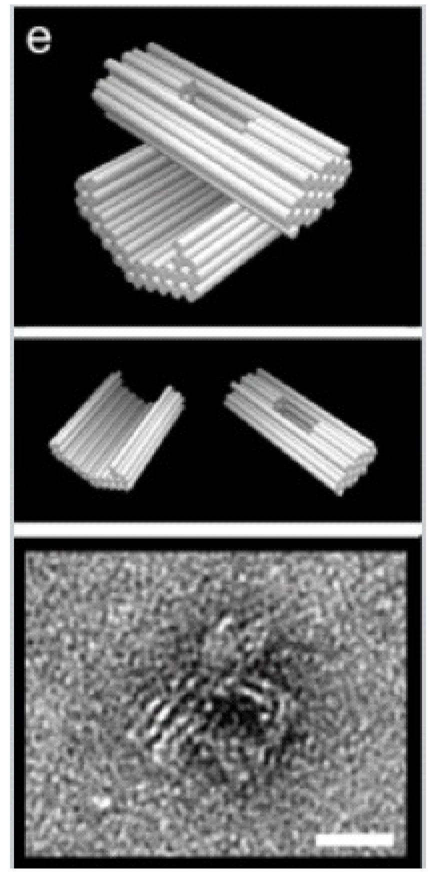
Folding DNA into Twisted and Curved Nanoscale Shapes

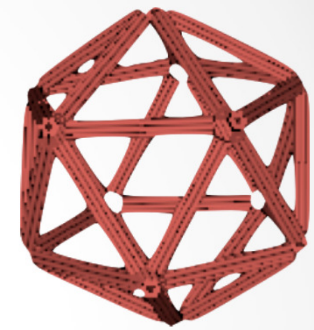
Hendrik Dietz, Shawn M. Douglas, & William M. Shih

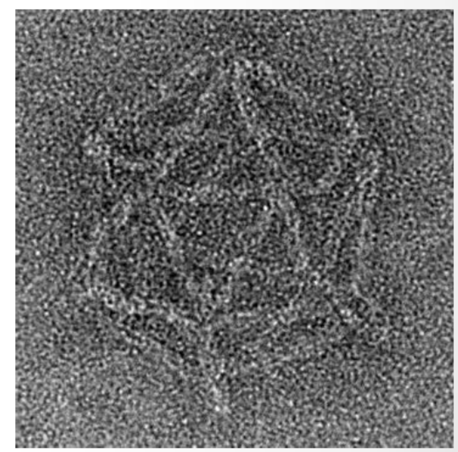
Science, 325:725–730, 7 August 2009.

CADnano







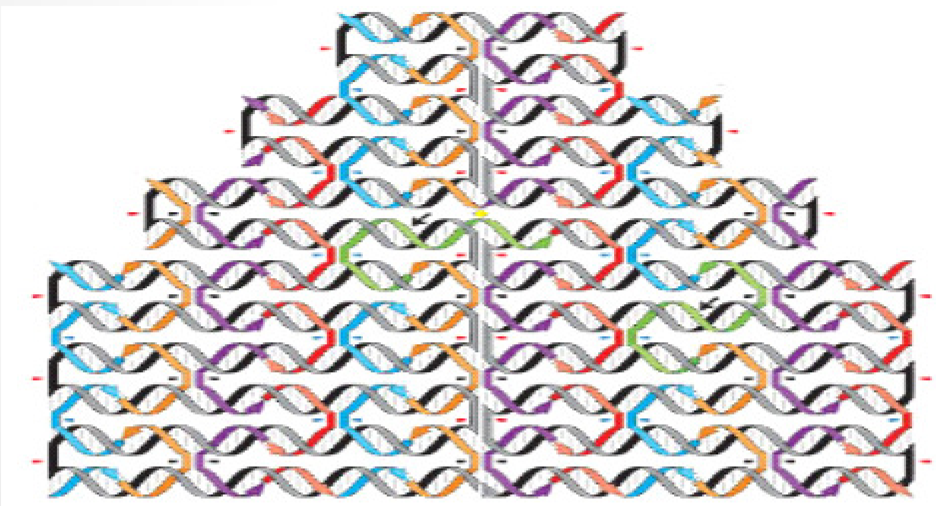


William Shih
Harvard

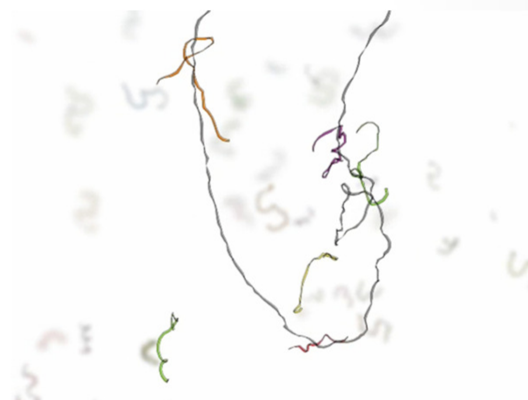
S.M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf and W. M. Shih
Self-assembly of DNA into nanoscale three-dimensional shapes, Nature (2009)

DNA Origami

- *Folding* long (7000bp) naturally occurring (viral) ssDNA
- By lots of short 'staple' strands that constrain it

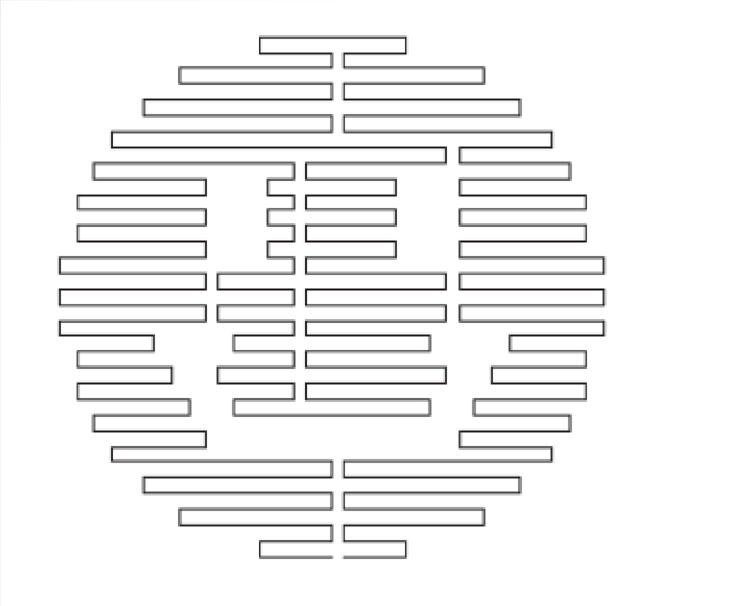


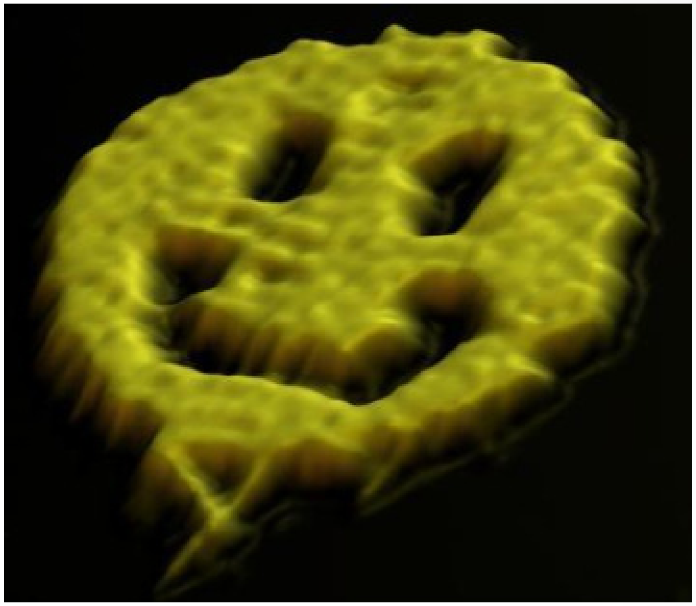
PWK Rothemund, *Nature* 440, 297 (2006)



Black: long viral strand
Color: short staple strands

DNA Origami



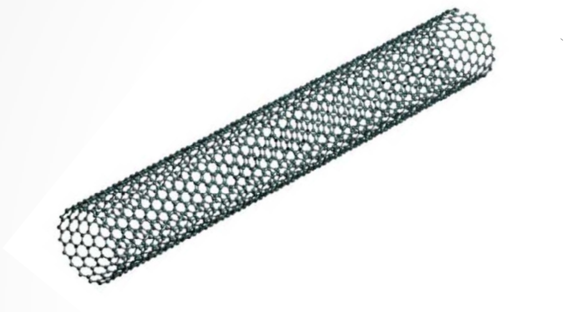


Paul Rothemund's "Disc with three holes" (2006)

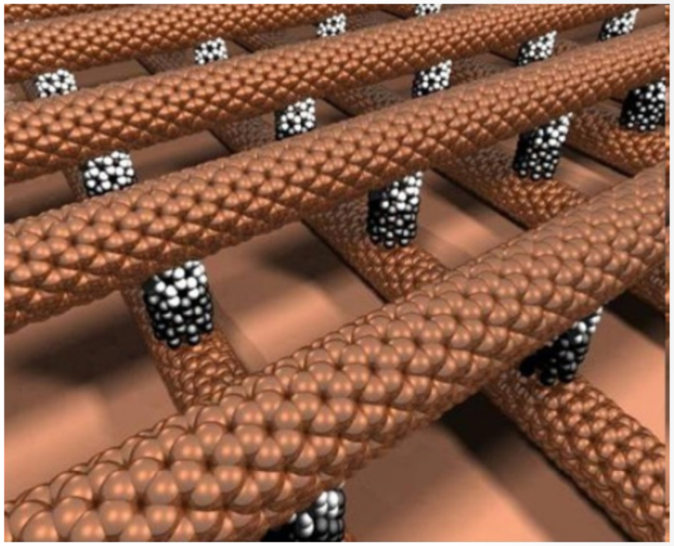
Paul W K Rothemund
California Institute of Technology

This means we can already self-assemble meso-scale structures.

DNA Circuit Boards

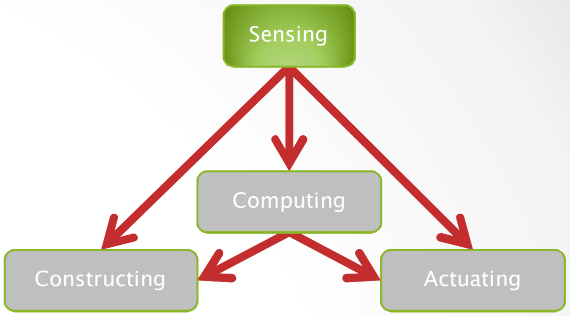


DNA-wrapped nanotubes



European Nanoelectronics Initiative Advisory Council

"What we are really making are tiny DNA circuit boards that will be used to assemble other components."
--Greg Wallraff, IBM

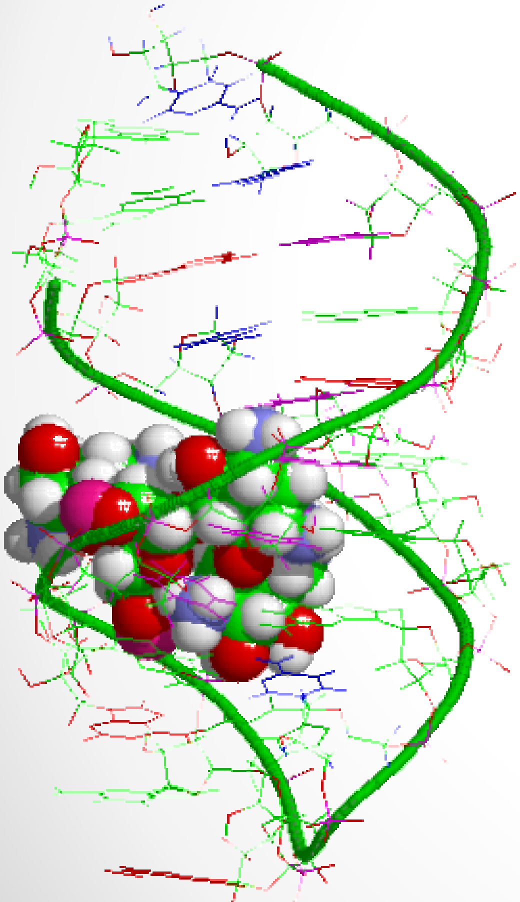


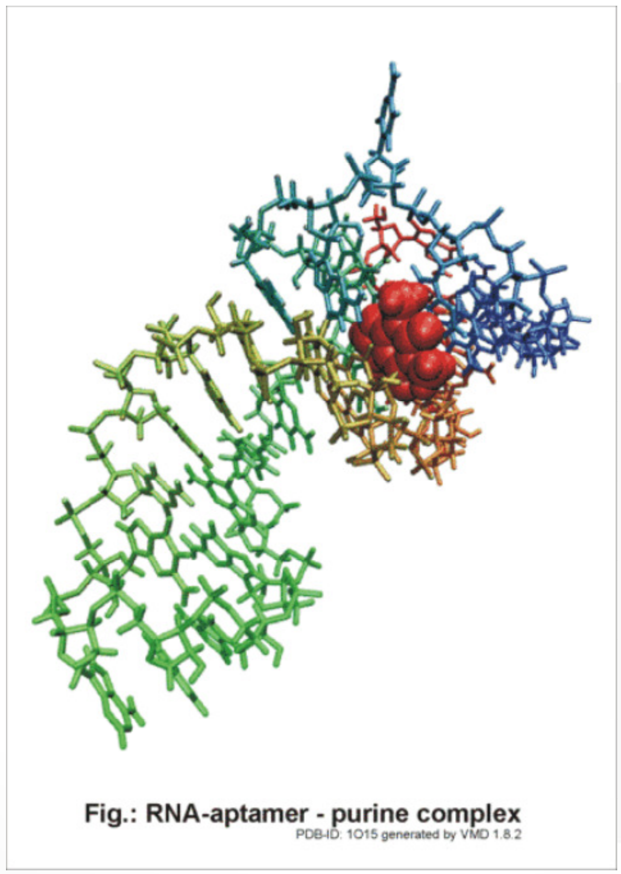
Sensing

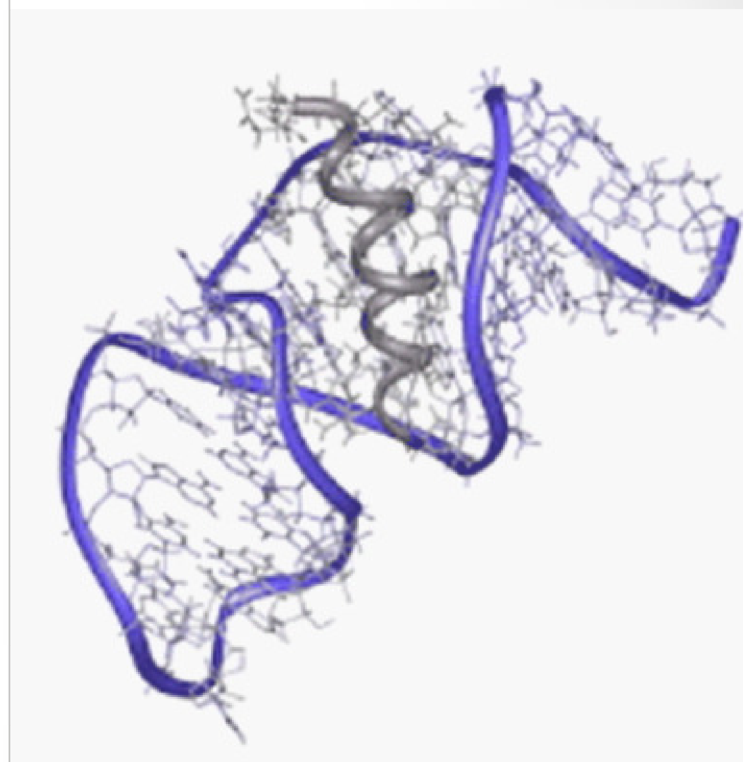
• • •

Aptamers

- Artificially eveloved DNA molecules that stick to anything you like (highly selectively).

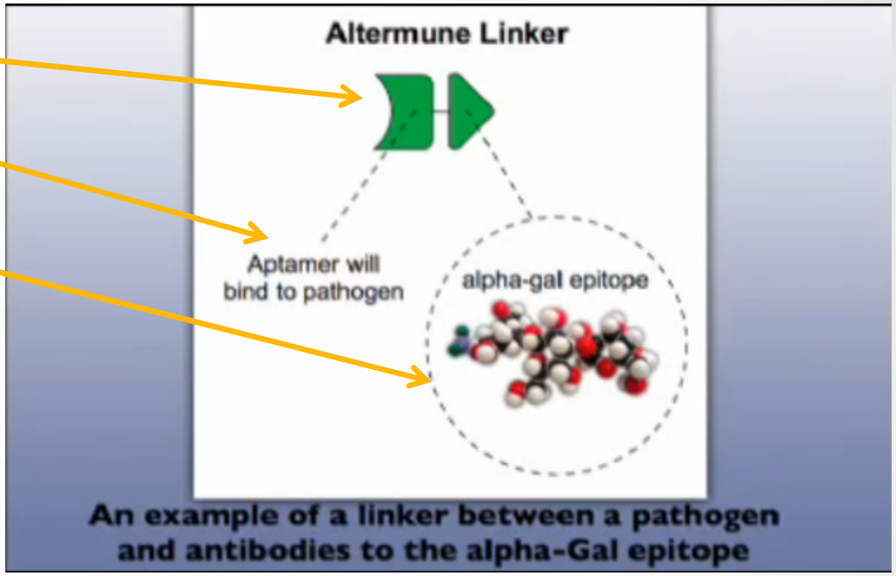




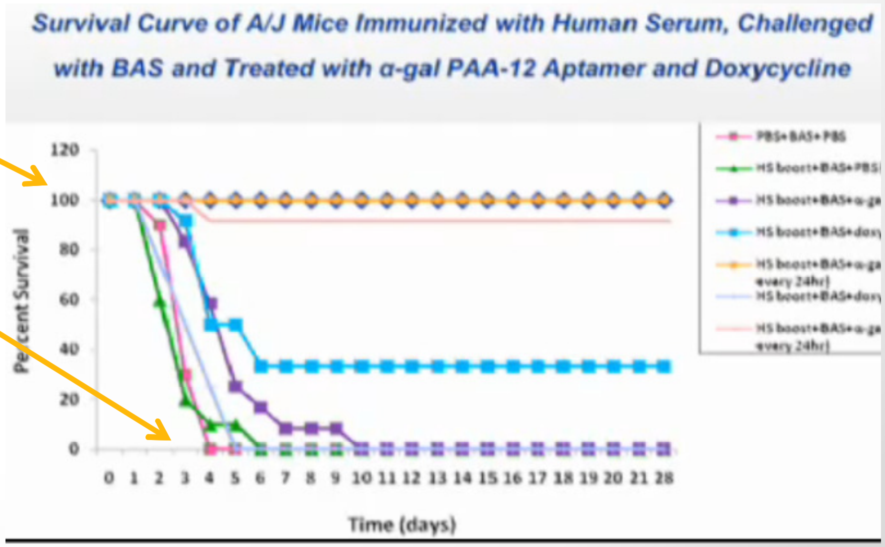


Pathogen Spotlights

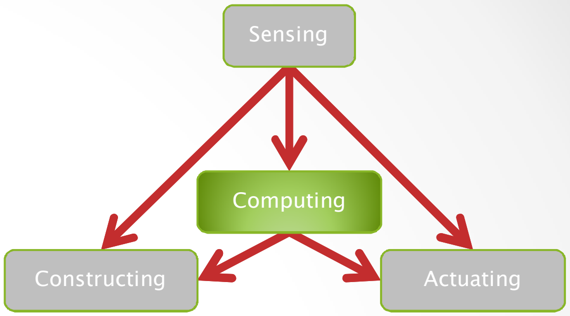
- DNA aptamer binds to:
  - A) a pathogen
  - B) a molecule our immune system already hates and immediately removes (eats) along with anything attached to it



- Result: instant immunity
  - Mice poisoned with Anthrax plus aptamer (100% survival)
  - Mice poinsoned with Anthrax (not so good)



Kary Mullis (incidentally, also Nobel prize for inventing the Polymerase Chain Reaction)

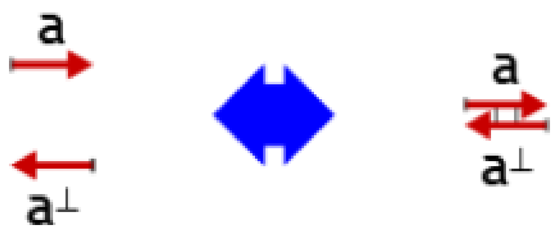


Computing

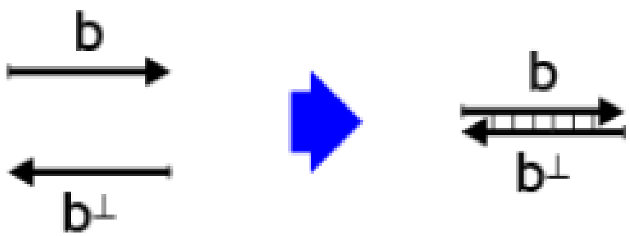
• • •

Rules of the Game

- Short complementary segments hybridize **reversibly**

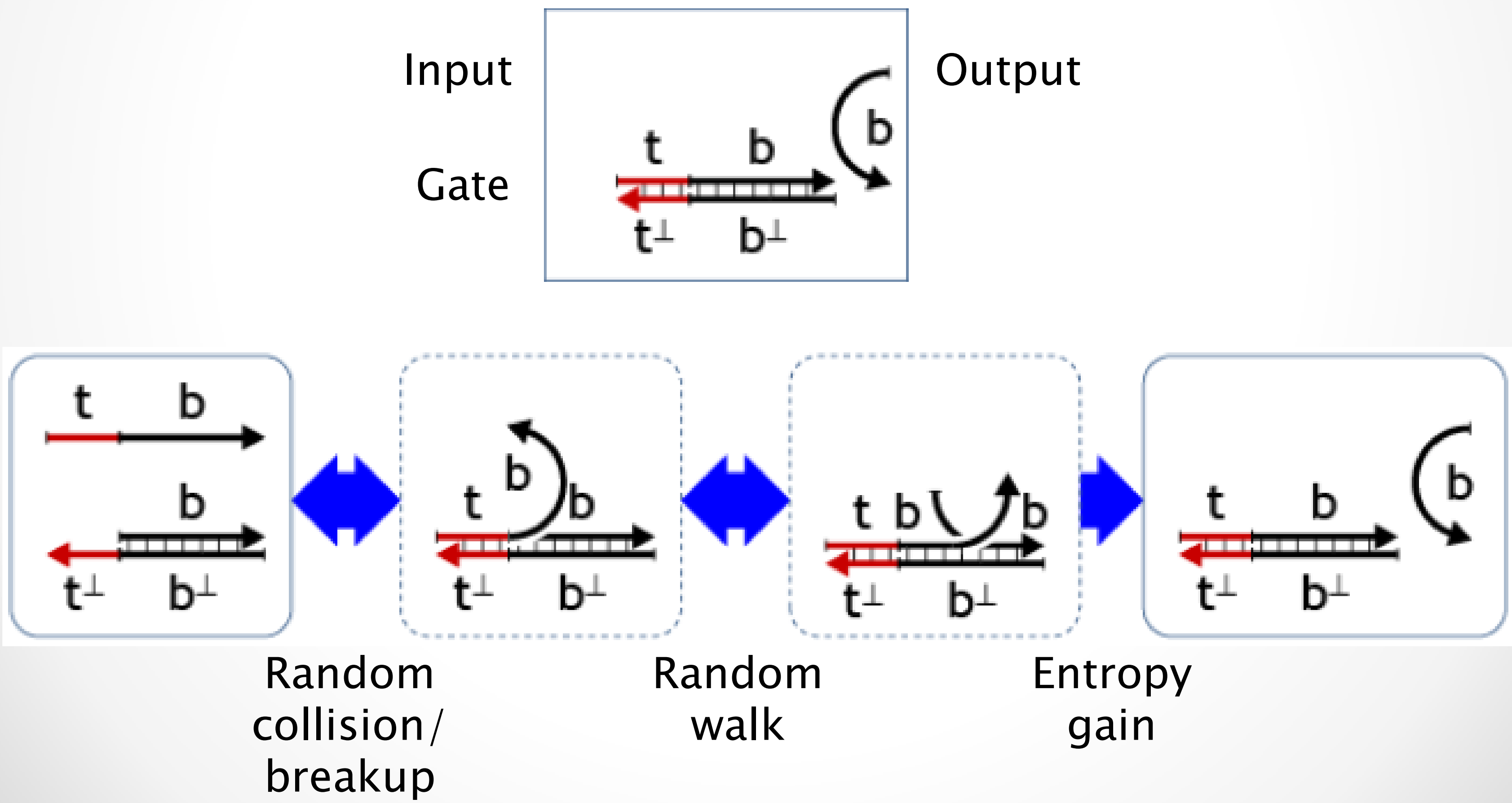


- Long complementary segments hybridize **irreversibly**



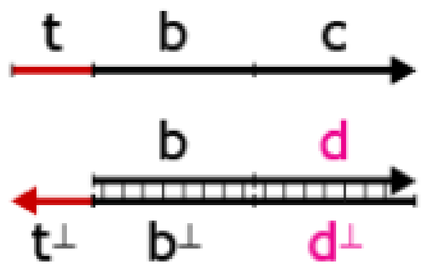
DNA Strand Displacement

- Short strand (**toehold**): reversible binding
- Long strand (**body**): irreversible binding

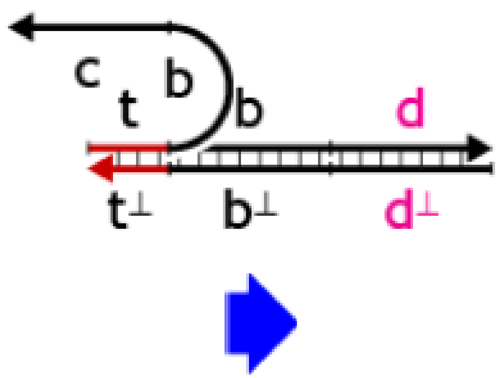


Failed Strand Displacement

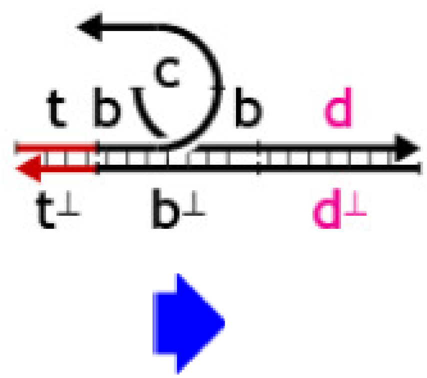
- What if the input does not match the gate?



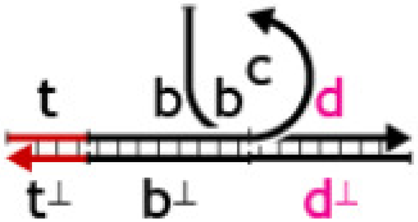
Failed Strand Displacement



Failed Strand Displacement

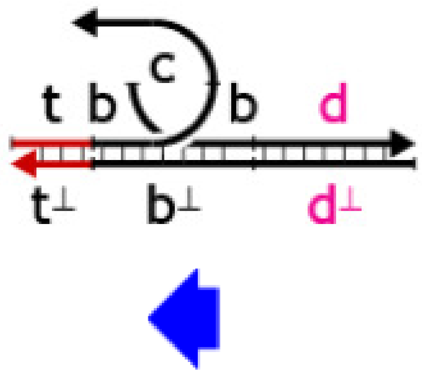


Failed Strand Displacement

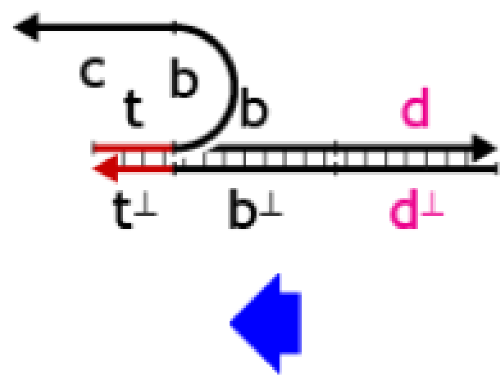


?

Failed Strand Displacement

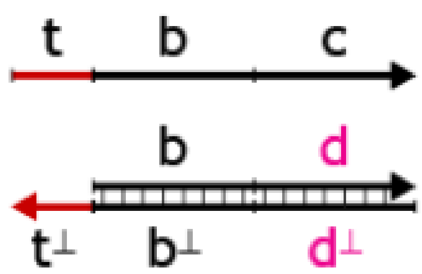


Failed Strand Displacement



Failed Strand Displacement

- Hence an incorrect binding will undo
  - That's why toeholds must bind reversibly



- Matching depends on the long segment only
  - Strand displacement succeeds iff the whole long segment matches
  - The address space is determined by the size of the long segment, which is unbounded (not by the size of the toehold)
  - The toehold is just a 'cache' of the address

What does DNA Compute?

- Electronics has *electrons*
  - All electrons are the same: you can only count them
  - *Few* electrons = False; *lots* of electrons = True
  - But Boolean Logic is only a necessary evil to build symbolic computation

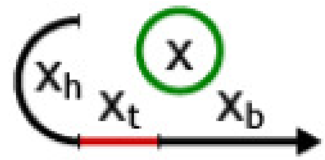
- DNA computing has *symbols* (DNA words)
  - DNA words are not all the same
  - Symbolic computation on abstract signals can be done *directly*
  - Signals are presented concurrently (in a soup)
  - No requirement to do Boolean Logic

- Then, what are our 'gates' (if not Boolean?)
  - Theory of Concurrency
  - Process Algebra as the "Boolean Algebra" of DNA Computing

Signals

- A signal is the representation of an abstract event
  - E.g. generated by a sensor
  - E.g. accepted by an effector
  - We are not limited to true/false

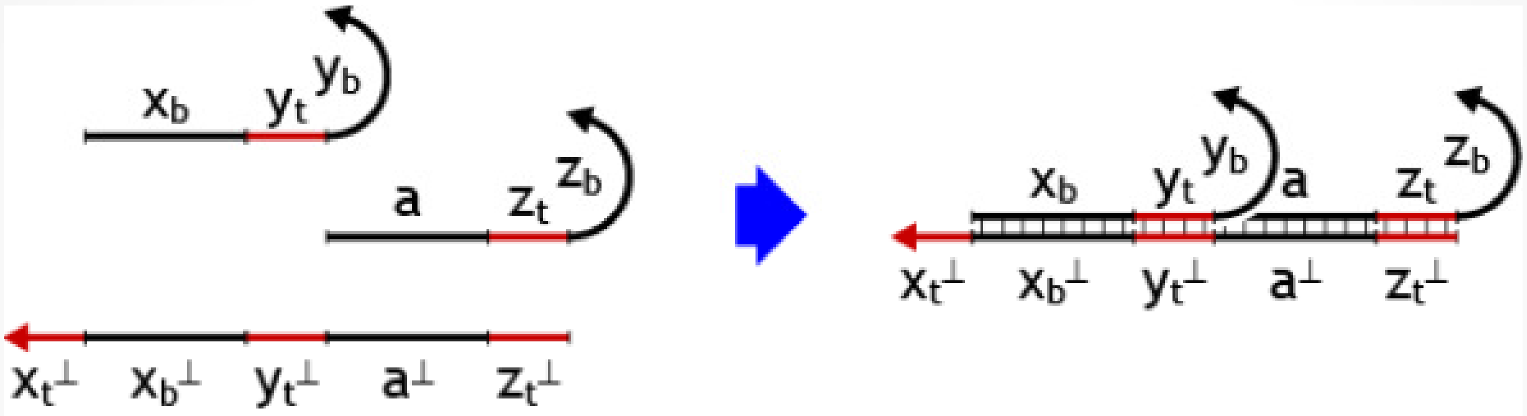
- 3-domain signals
  - x_h: hystory (ignore)
  - x_t: toehold (binding)
  - x_b: body (recognition)



- Signals (single stranded DNA) are prepared by (artificial) DNA synthesis

Gates

- Double-stranded structures with free toeholds



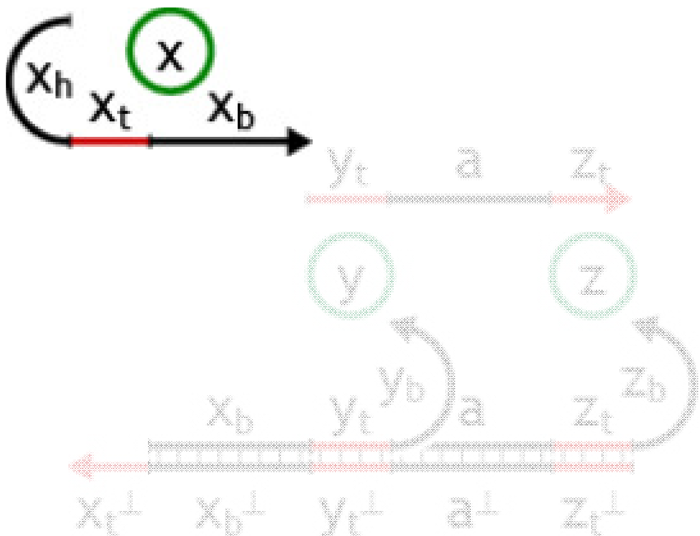
- Gates are prepared by self-assembly from single-stranded DNA that is synthesized

Fork Gate

- $x \rightarrow y + z$

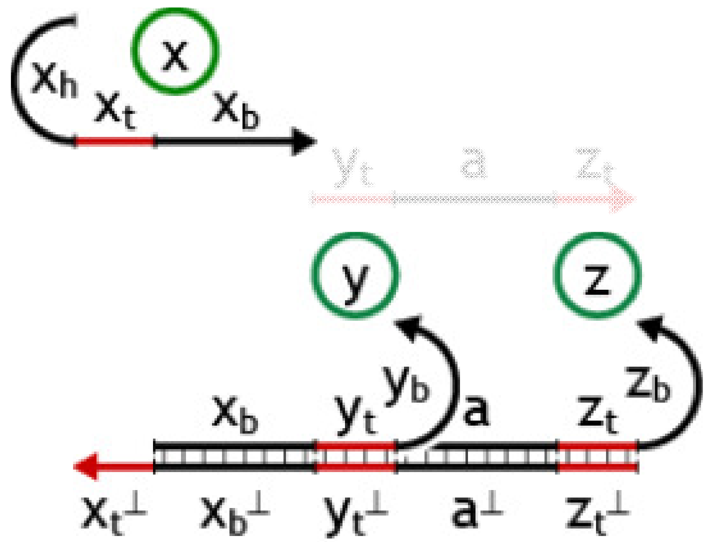
- $x \rightarrow y + 0$ transform x to y (transducer)
- $x \rightarrow x + y$ linear production of y (catalyst)
- $x \rightarrow x + x$ exponential production of x (amplifier)

Fork Gate



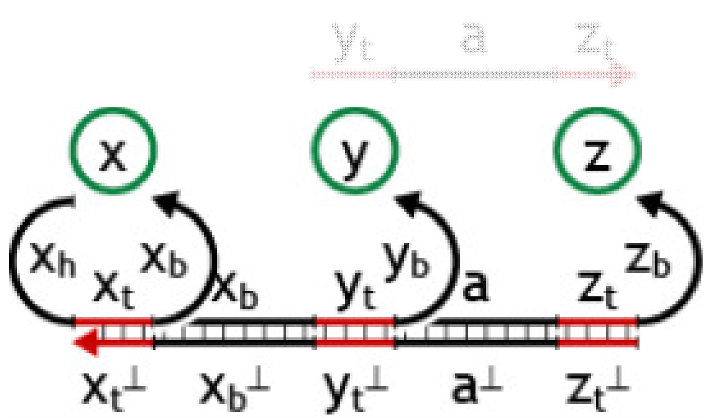
This is the Fork Gate structure

Fork Gate

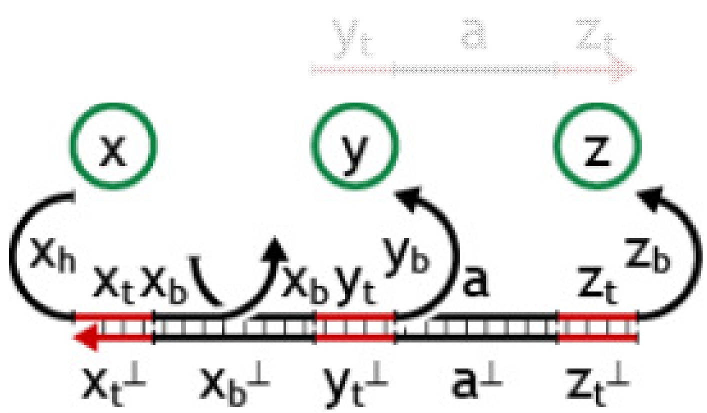


x
x_h
x_t
x_b
y_t
a
z_t
y
z
y_b
z_b
x_b
y_t
a
z_t
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$a^\perp$
$z_t^\perp$

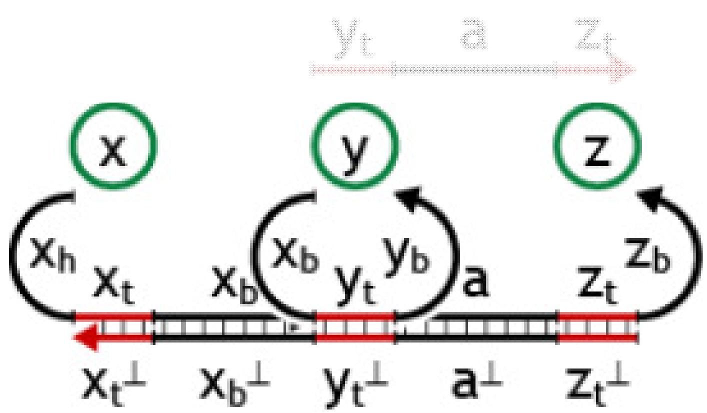
Fork Gate



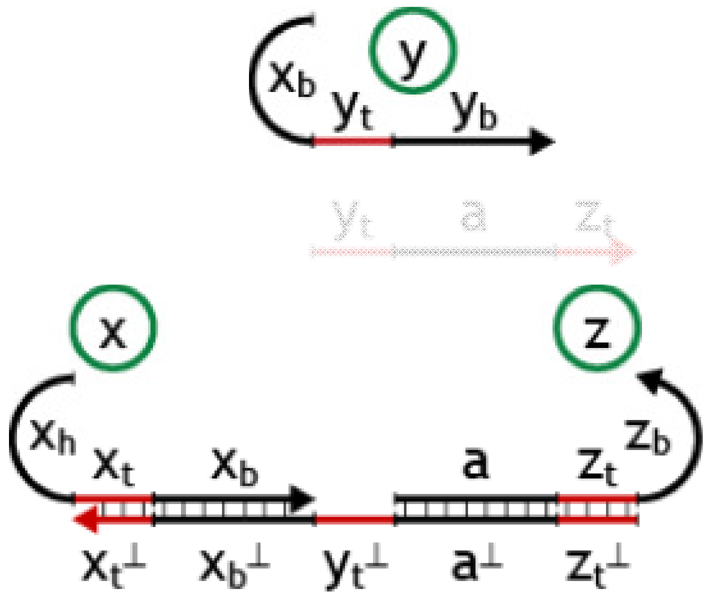
Fork Gate



Fork Gate

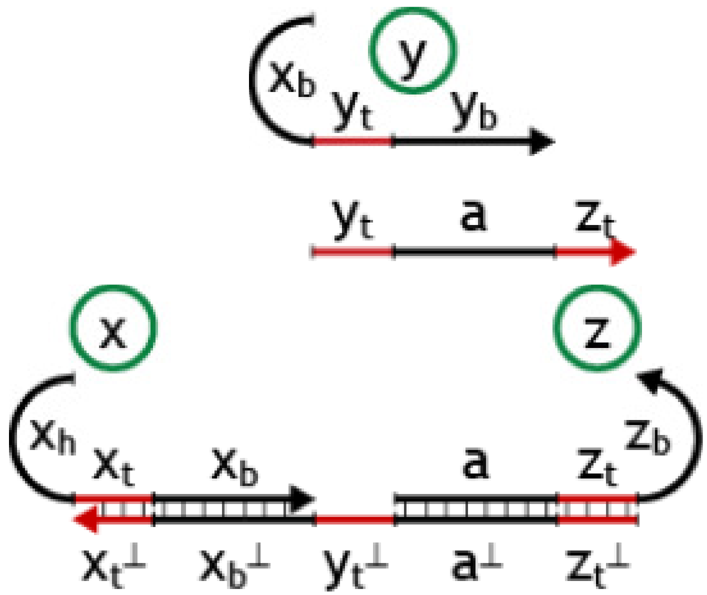


Fork Gate

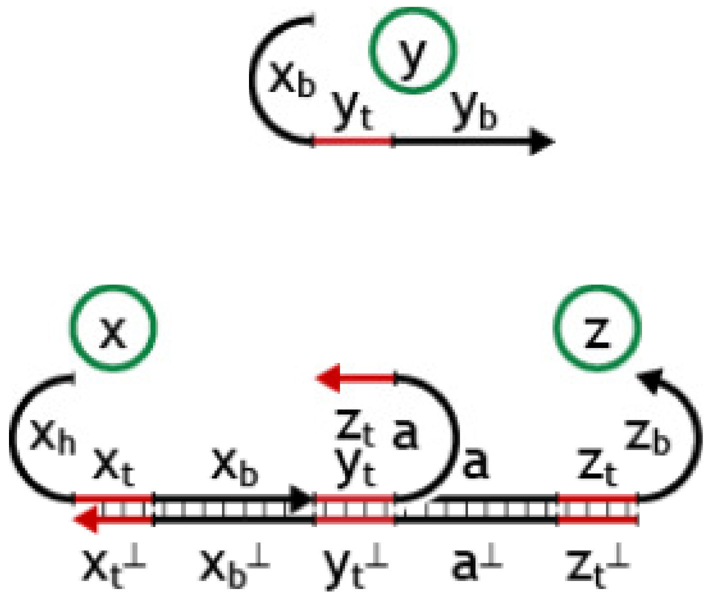


y
x_b
y_t
y_b
y_t
a
z_t
x
z
x_h
x_t
x_b
a
z_t
z_b
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$a^\perp$
$z_t^\perp$

Fork Gate

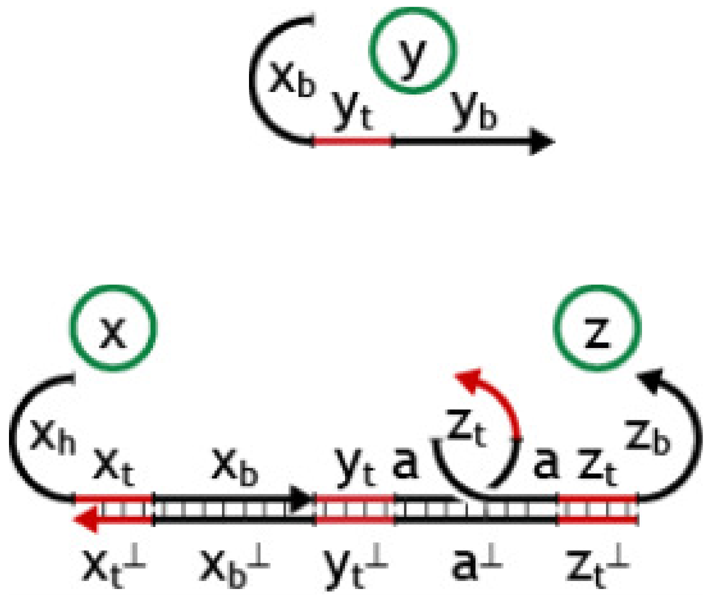

y
x_b
y_t
y_b
y_t
a
z_t
x
z
x_h
x_t
x_b
a
z_t
z_b
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$a^\perp$
$z_t^\perp$

Fork Gate

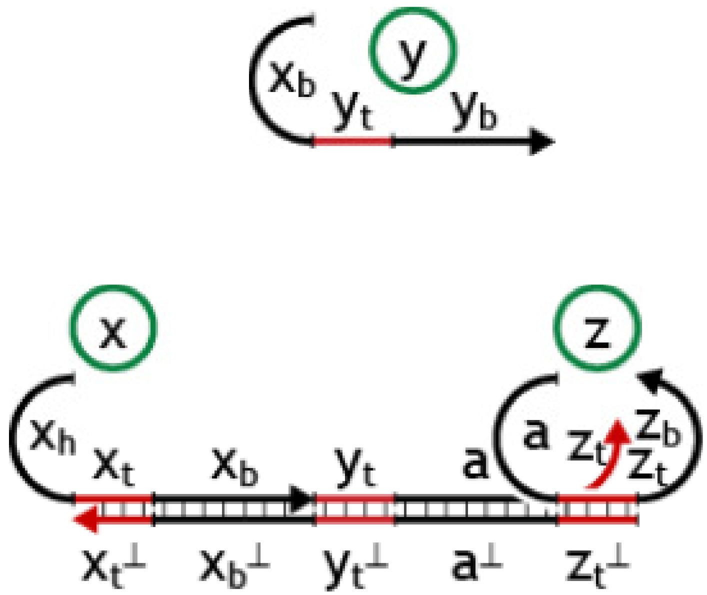


y
x_b
y_t
y_b
x
z
x_h
x_t
x_b
z_t
a
y_t
a
z_t
z_b
$x_t^{\perp}$
$x_b^{\perp}$
$y_t^{\perp}$
$a^{\perp}$
$z_t^{\perp}$

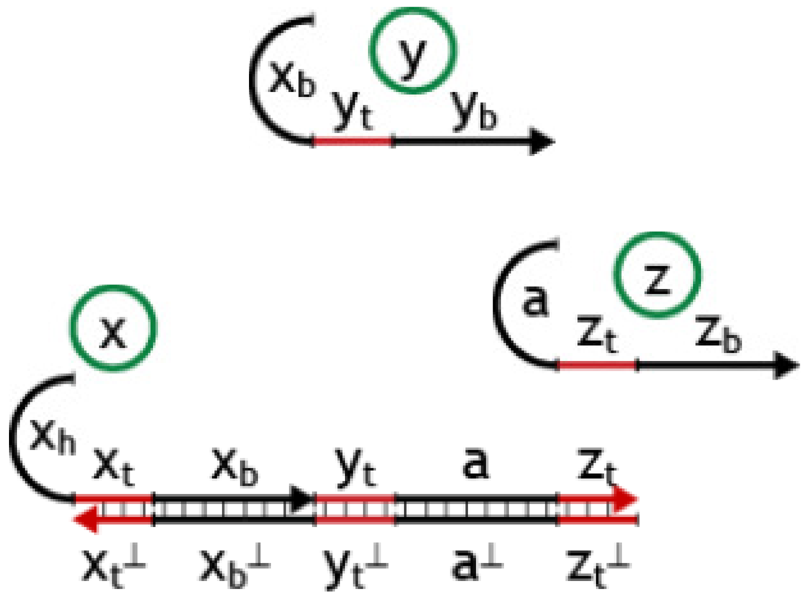
Fork Gate



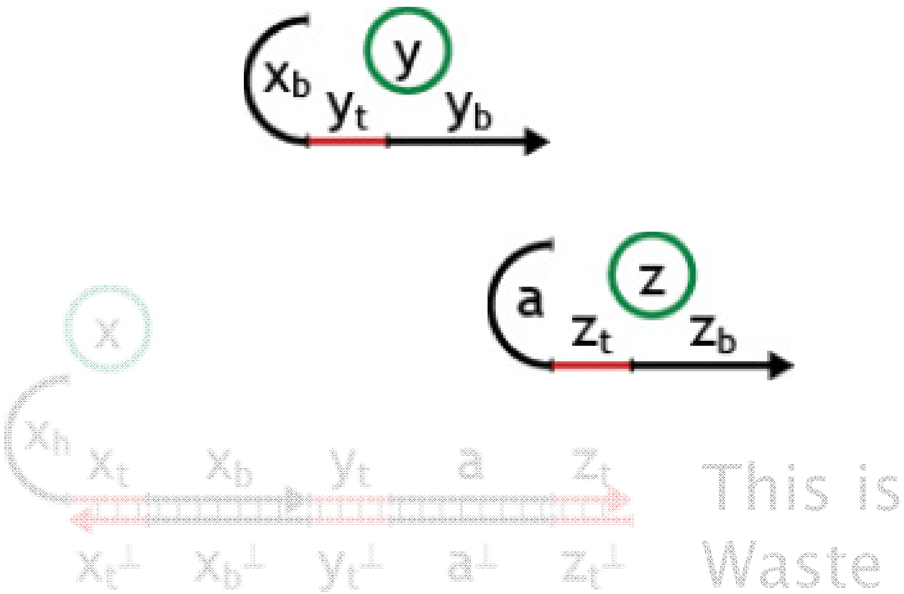
Fork Gate


y
x_b
y_t
y_b
x
z
x_h
x_t
x_b
y_t
a
a
z_t
z_b
z_t
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$a^\perp$
$z_t^\perp$

Fork Gate



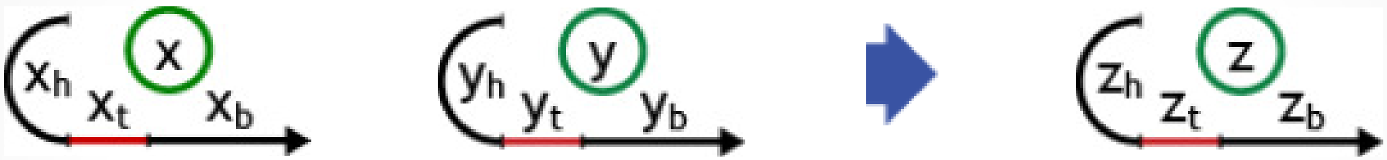
Fork Gate



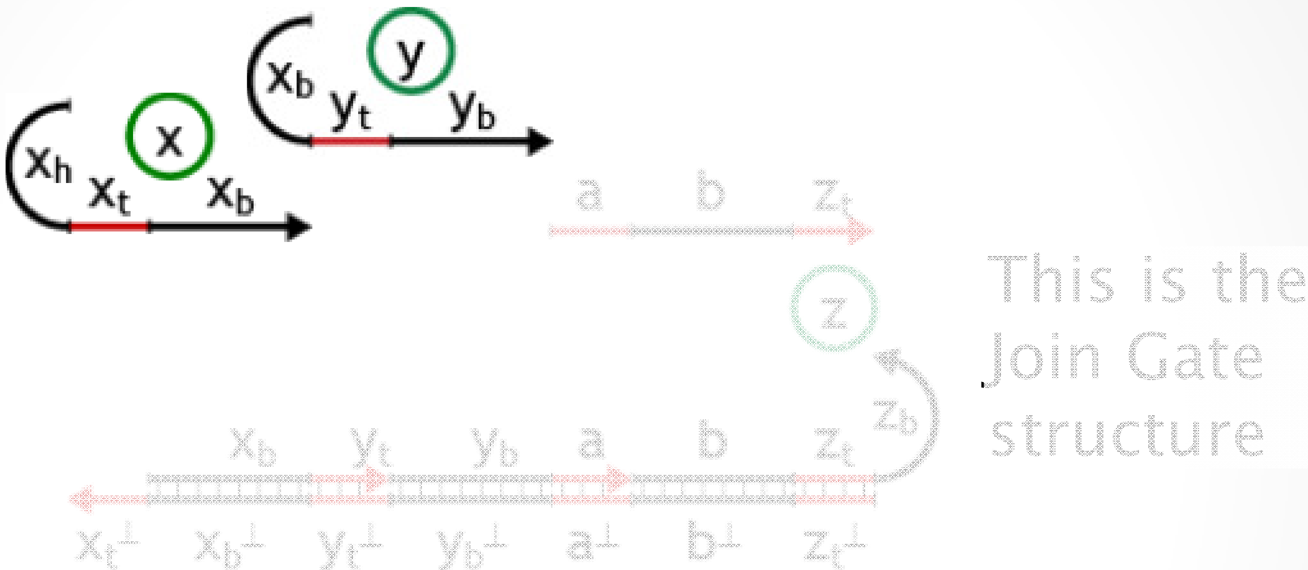
y
x_b
y_t
y_b
z
a
z_t
z_b
x
x_h
x_t x_b y_t a z_t
$x_t^\perp$ $x_b^\perp$ $y_t^\perp$ $a^\perp$ $z_t^\perp$
This is Waste

Join Gate

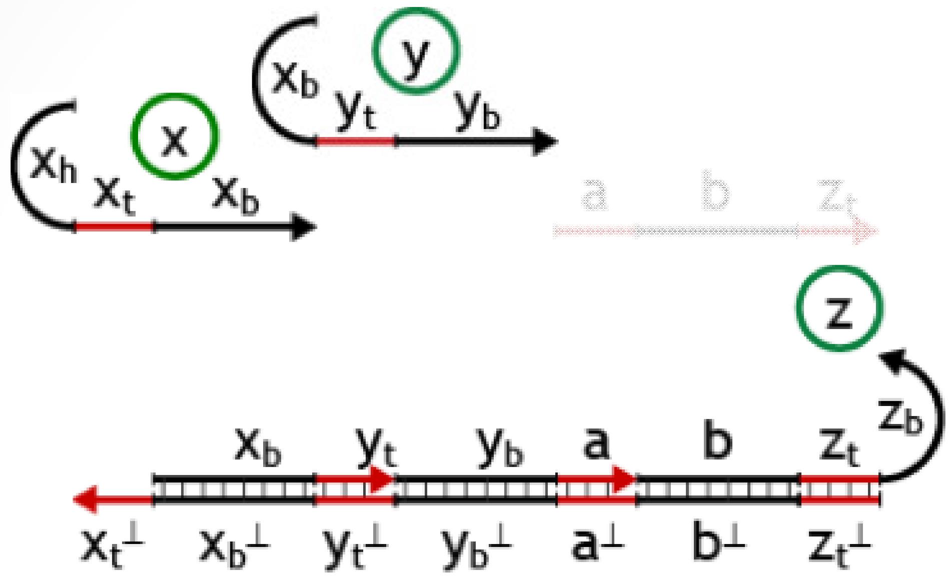
- $x + y \rightarrow z$



Join Gate

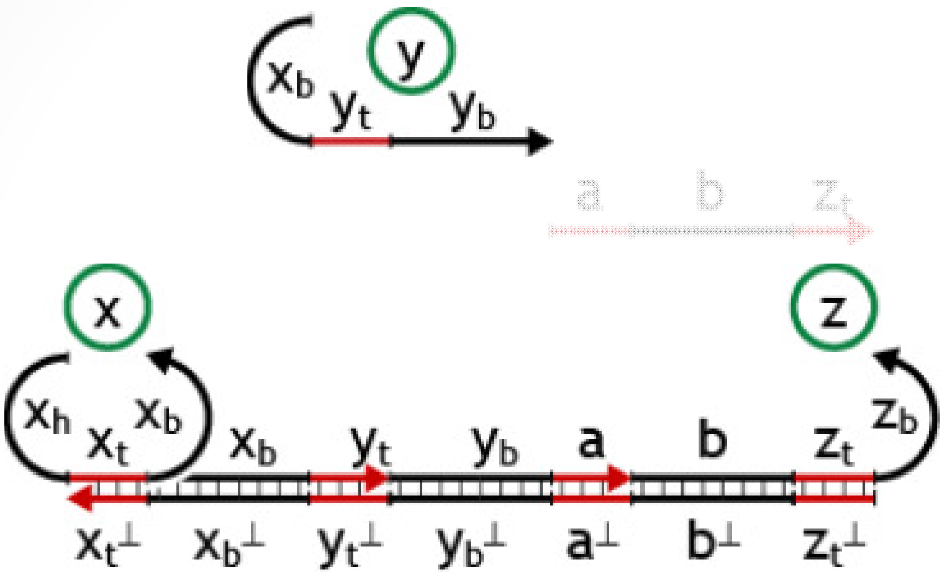


Join Gate



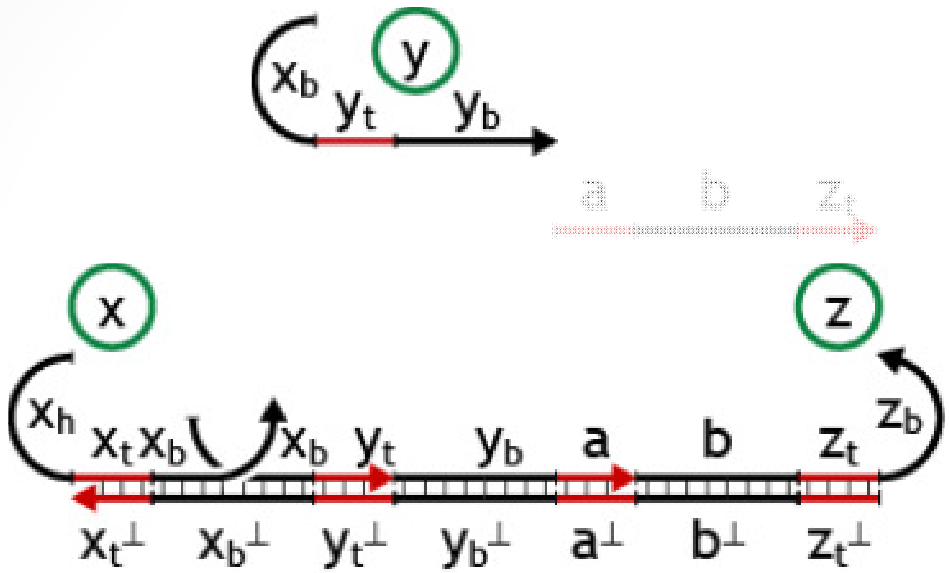
x_h
x_t
x
x_b
x_b
y_t
y
y_b
a
b
z_t
z
z_b
x_b
y_t
y_b
a
b
z_t
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$y_b^\perp$
$a^\perp$
$b^\perp$
$z_t^\perp$

Join Gate

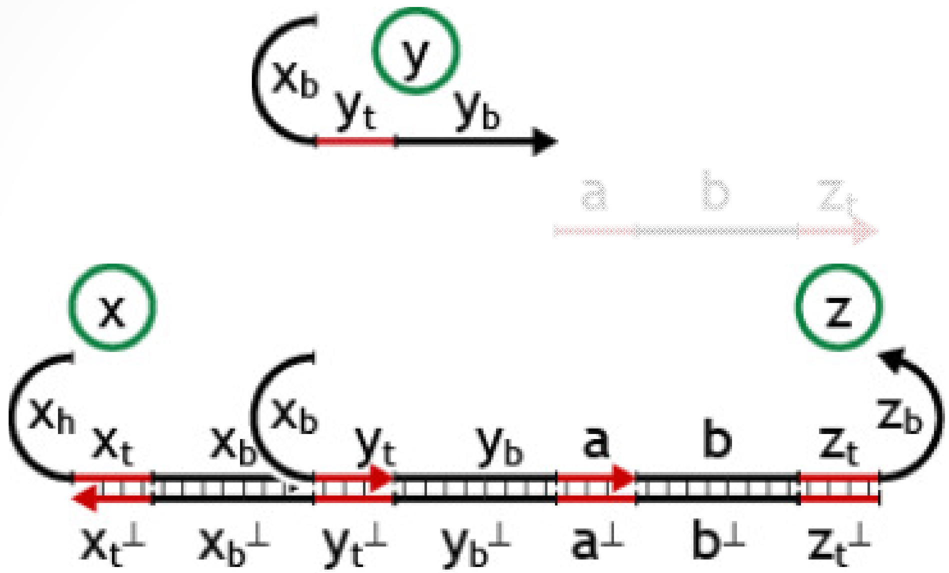


y
x_b y_t y_b
a b z_t
x
z
x_h x_t x_b
x_b y_t y_b a b z_t z_b
$x_t^\perp$ $x_b^\perp$ $y_t^\perp$ $y_b^\perp$ $a^\perp$ $b^\perp$ $z_t^\perp$

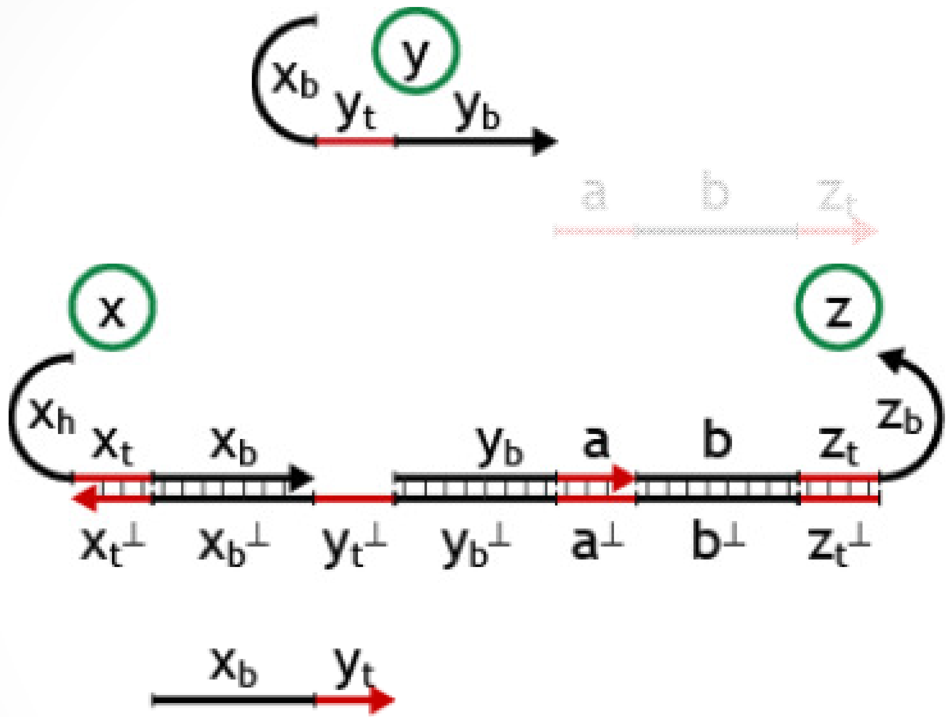
Join Gate



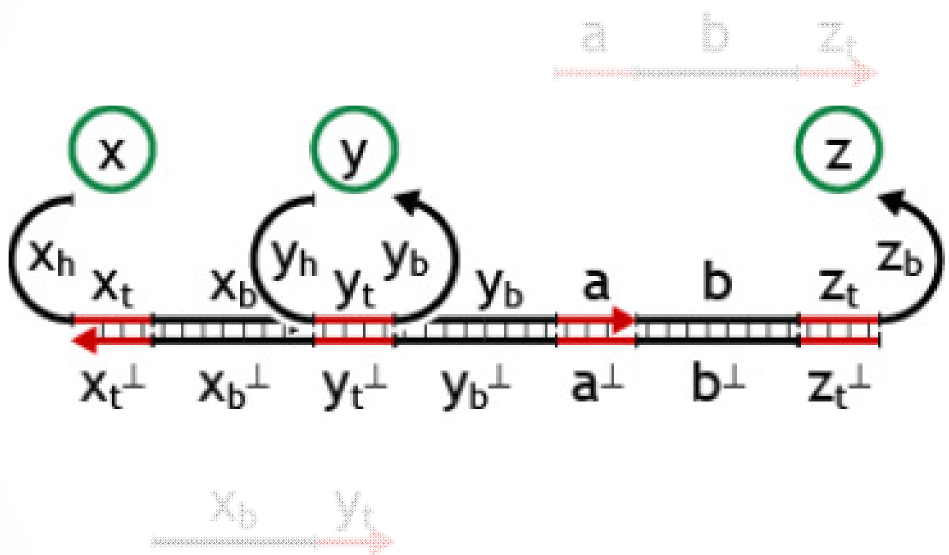
Join Gate



Join Gate

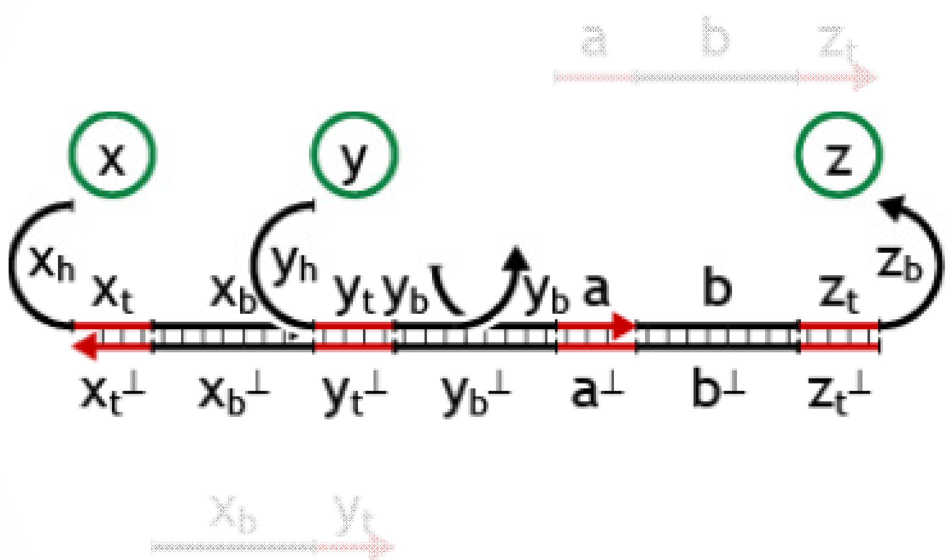


Join Gate



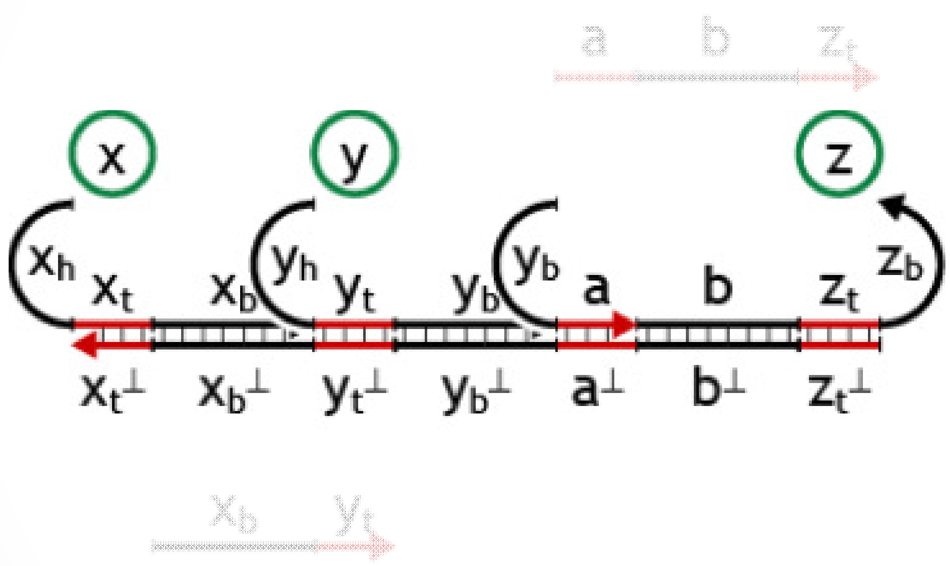
a b z_t
x
y
z
x_h x_t x_b y_h y_t y_b y_b a b z_t z_b
$x_t^\perp$ $x_b^\perp$ $y_t^\perp$ $y_b^\perp$ $a^\perp$ $b^\perp$ $z_t^\perp$
x_b y_t

Join Gate



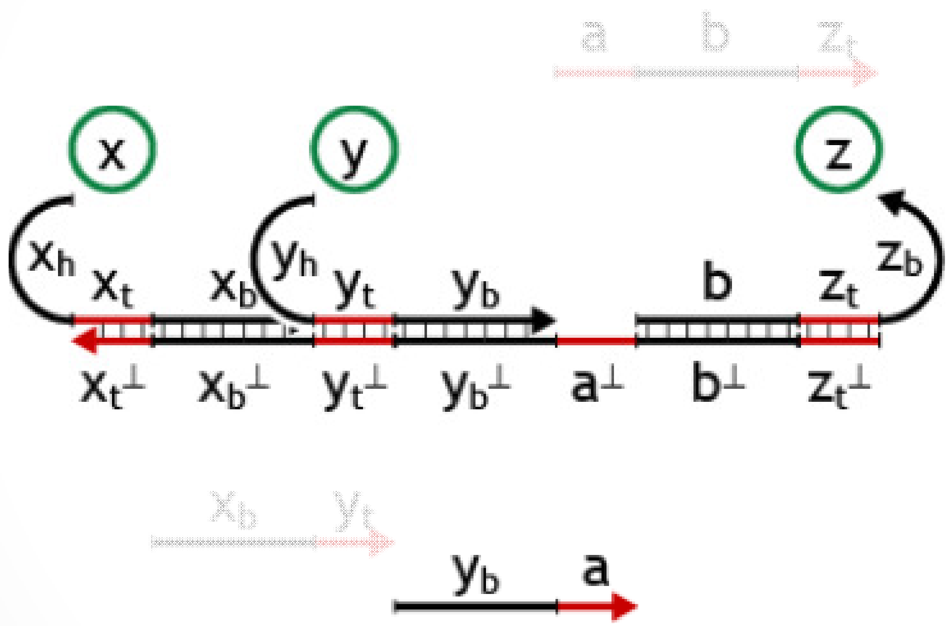
a
b
z_t
x
y
z
x_h
x_t
x_b
y_h
y_t
y_b
y_b
a
b
z_t
z_b
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$y_b^\perp$
$a^\perp$
$b^\perp$
$z_t^\perp$
x_b
y_t

Join Gate

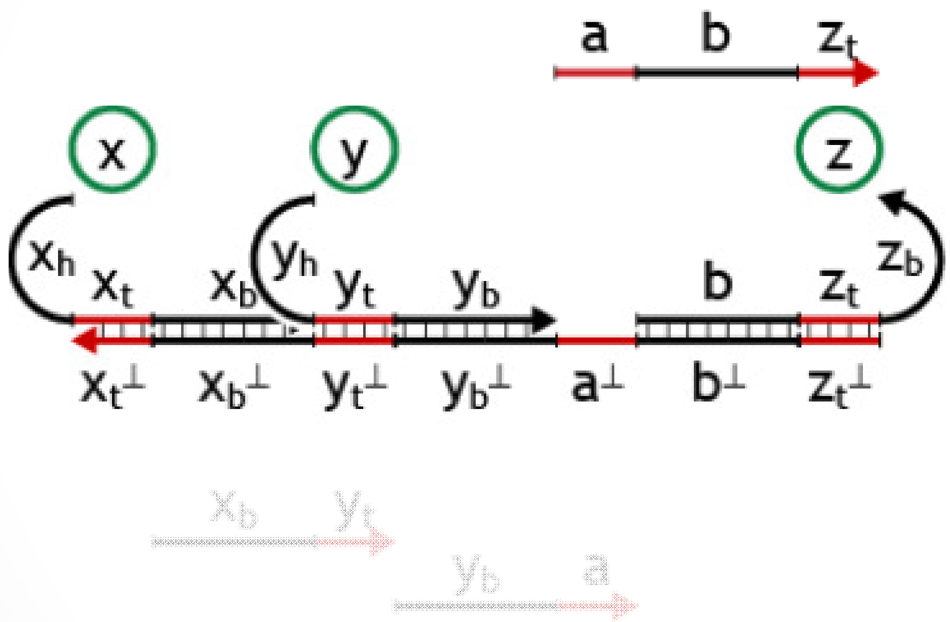


a
b
z_t
x
y
z
x_h
x_t
x_b
y_h
y_t
y_b
y_b
a
b
z_t
z_b
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$y_b^\perp$
$a^\perp$
$b^\perp$
$z_t^\perp$
x_b
y_t

Join Gate

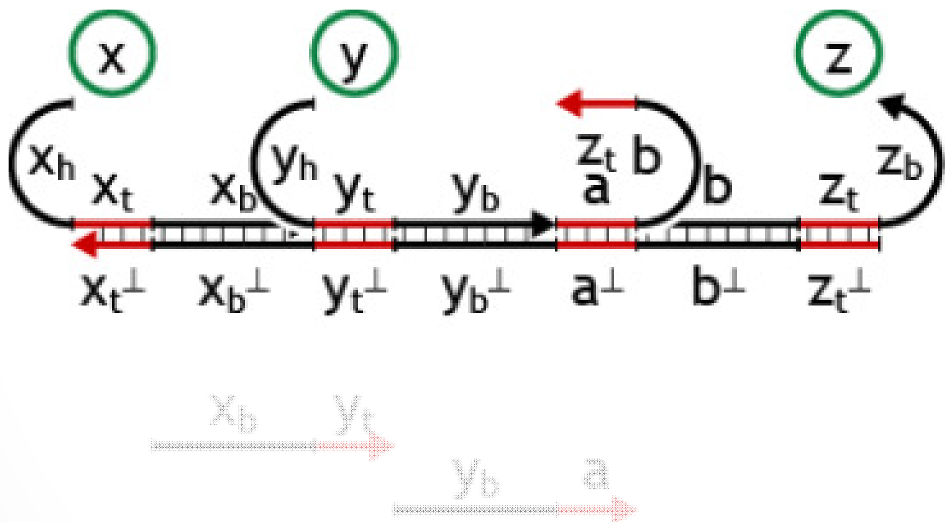


Join Gate



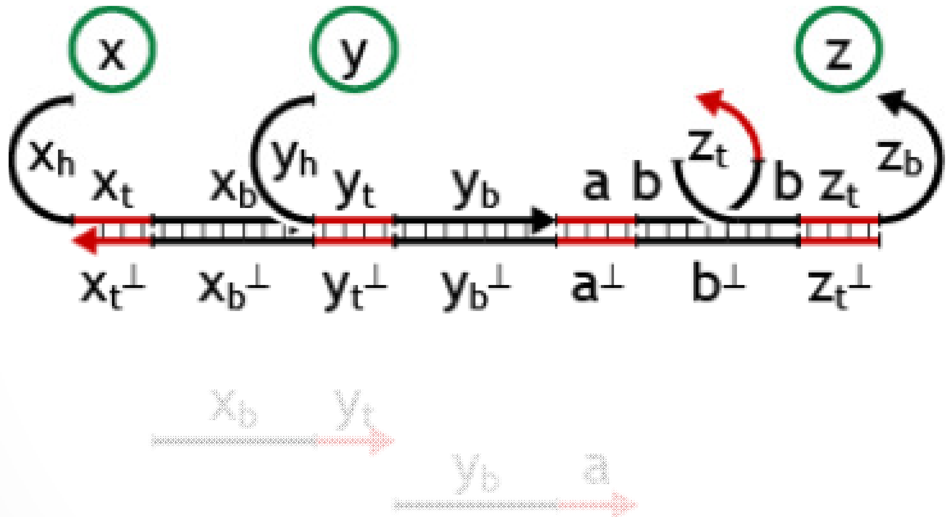
a
b
z_t
x
y
z
x_h
x_t
x_b
y_h
y_t
y_b
b
z_t
z_b
$x_t^{\perp}$
$x_b^{\perp}$
$y_t^{\perp}$
$y_b^{\perp}$
$a^{\perp}$
$b^{\perp}$
$z_t^{\perp}$
x_b
y_t
y_b
a

Join Gate

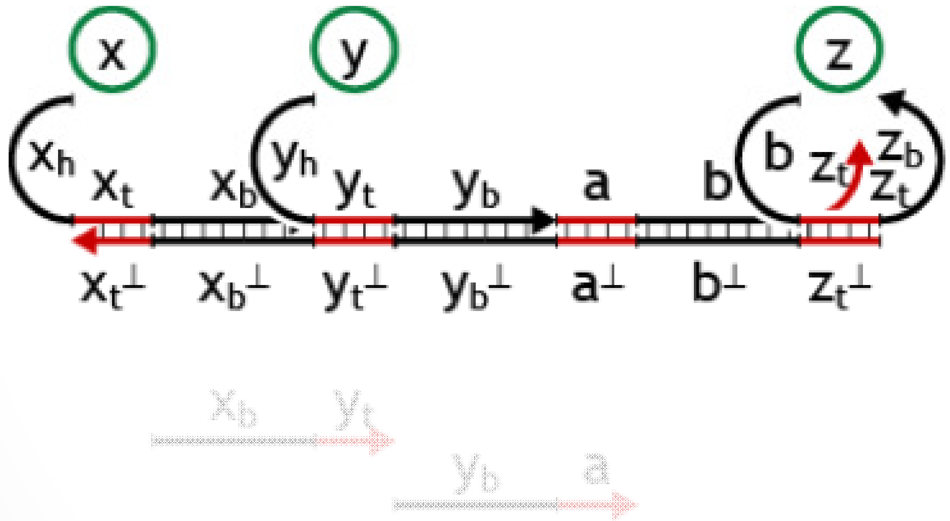


x
y
z
x_h
x_t
x_b
y_h
y_t
y_b
z_t b
a
b
z_t
z_b
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$y_b^\perp$
$a^\perp$
$b^\perp$
$z_t^\perp$
x_b
y_t
y_b
a

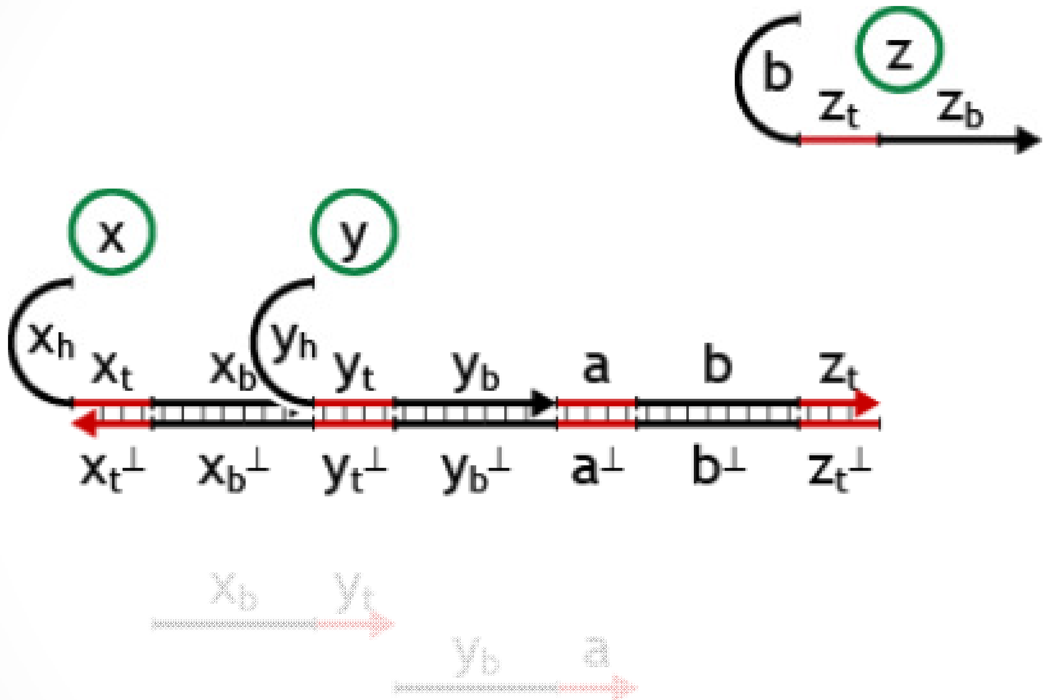
Join Gate


x
y
z
x_h
x_t
x_b
y_h
y_t
y_b
a
b
z_t
b
z_t
z_b
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$y_b^\perp$
$a^\perp$
$b^\perp$
$z_t^\perp$
x_b
y_t
y_b
a

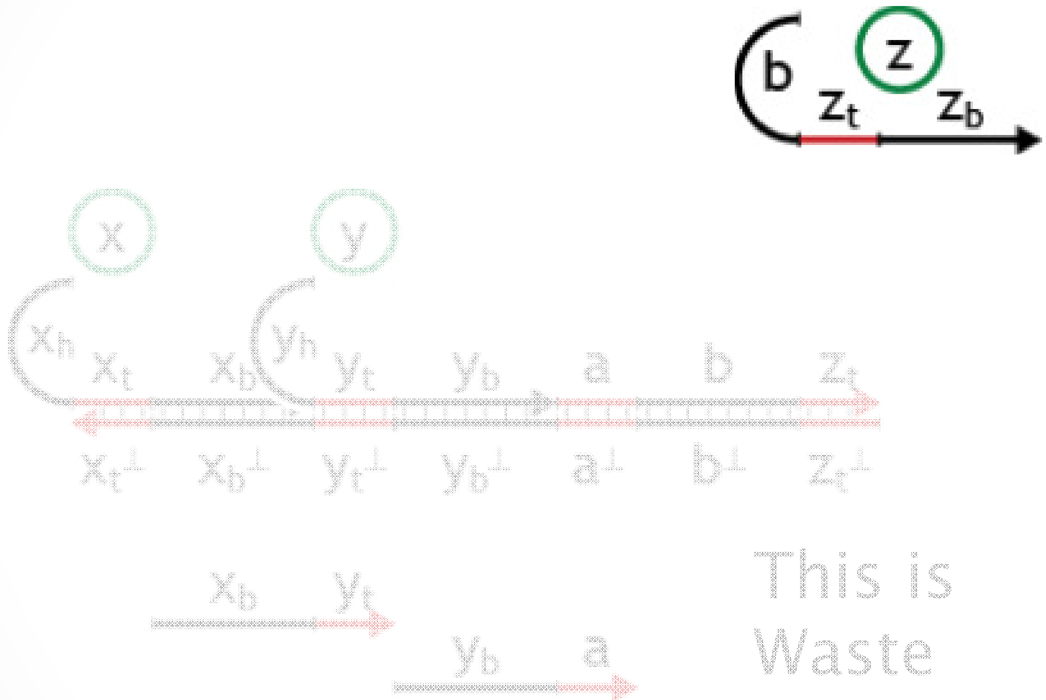
Join Gate



Join Gate



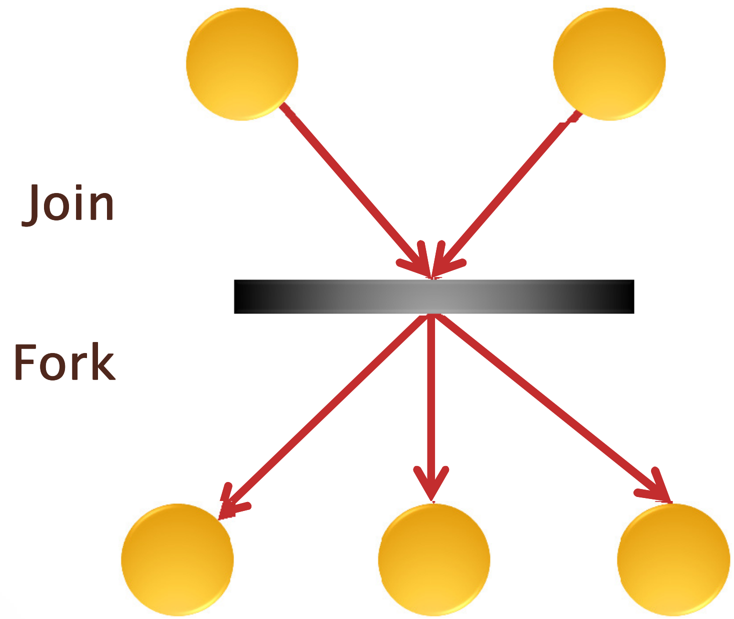
Join Gate



b
z
z_t
z_b
x
y
x_h
x_t
x_b
y_h
y_t
y_b
a
b
z_t
$x_t^\perp$
$x_b^\perp$
$y_t^\perp$
$y_b^\perp$
$a^\perp$
$b^\perp$
$z_t^\perp$
x_b
y_t
y_b
a
This is Waste

Strand Algebra

- Join + Fork + Populations = Petri Nets

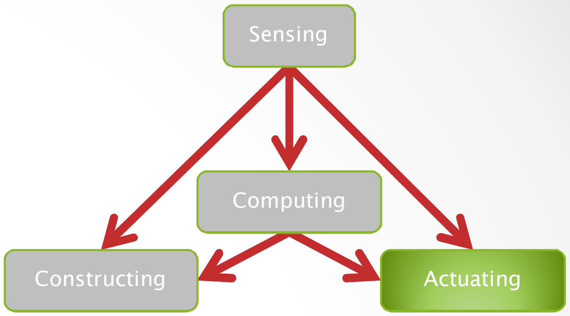


Gate Design Verification

- Active garbage
  - The active join residuals slow down the performance of following joins.
  - → Add a garbage collector to remove the active residuals.

- Interference between gates
  - The join garbage collector interferes with the fork gate.
  - → Modify the fork gate to remove the interference.

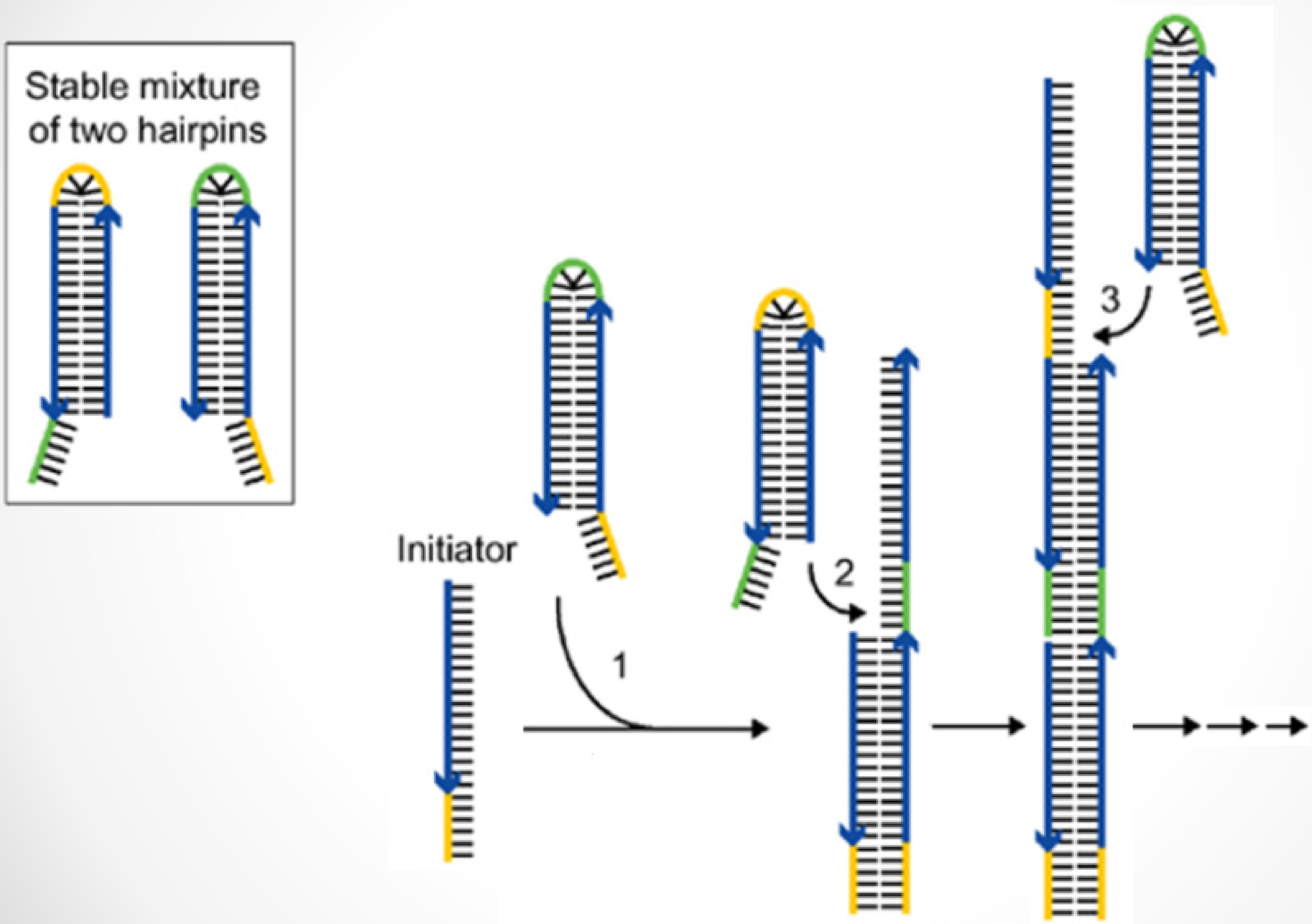
- What else could go wrong?
  - Endless possibilities.
  - → Prove that the fork/join gate structures correctly implement fork/join in all larger circuits.



Actuating

...

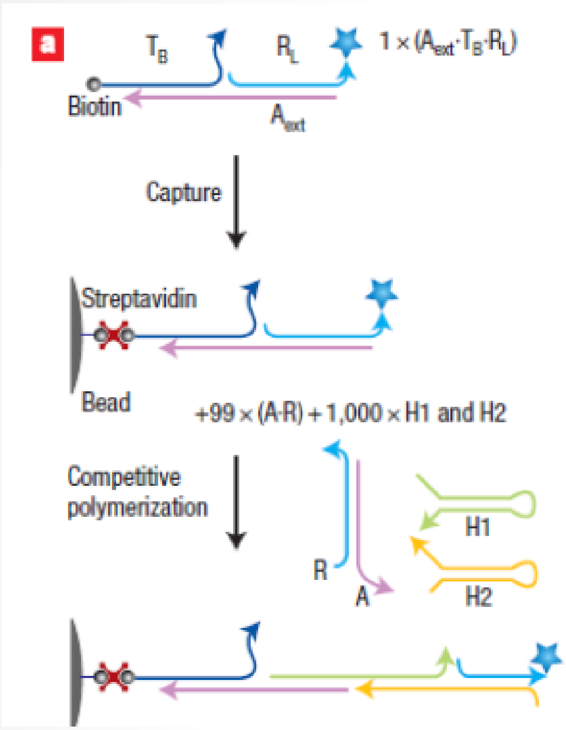
Hybridization Chain Reaction



n by hybridization
chain reaction

Robert M. Dirks† and Niles A. Pierce‡,§

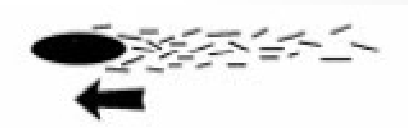
Polymerization Motor

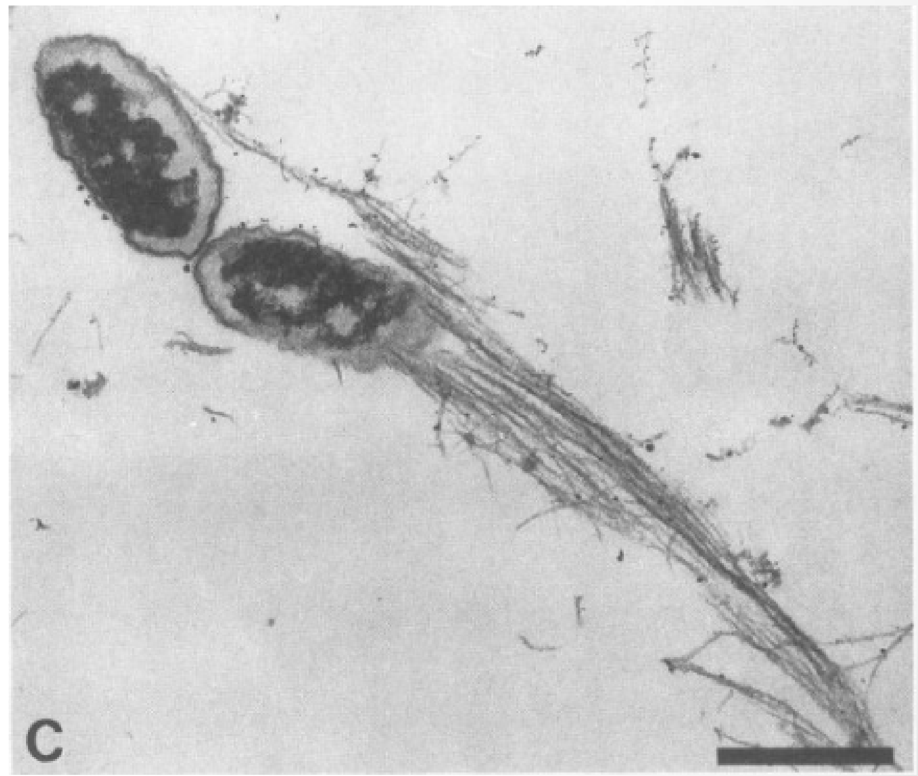


An autonomous polymerization motor powered by DNA hybridization

SUVIR VENKATARAMAN[1], ROBERT M. DIRKS[1], PAUL W. K. ROTHEMUND[2,3], ERIK WINFREE[2,3] AND NILES A. PIERCE[1,4]*

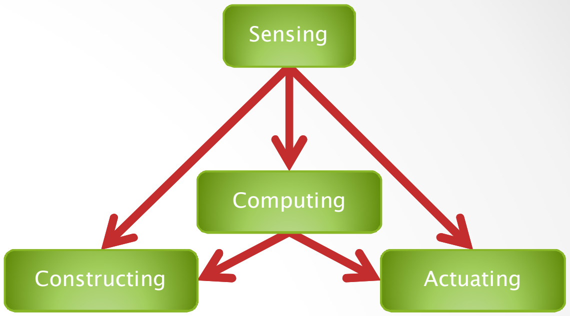
Rickettsia (spotted fever)





Directional Actin Polymerization Associated with Spotted Fever Group Rickettsia Infection of Vero Cells

ROBERT A. HEINZEN, STANLEY F. HAYES, MARIUS G. PEACOCK, AND TED HACKSTADT*



Curing

• • •

A Doctor in Each Cell

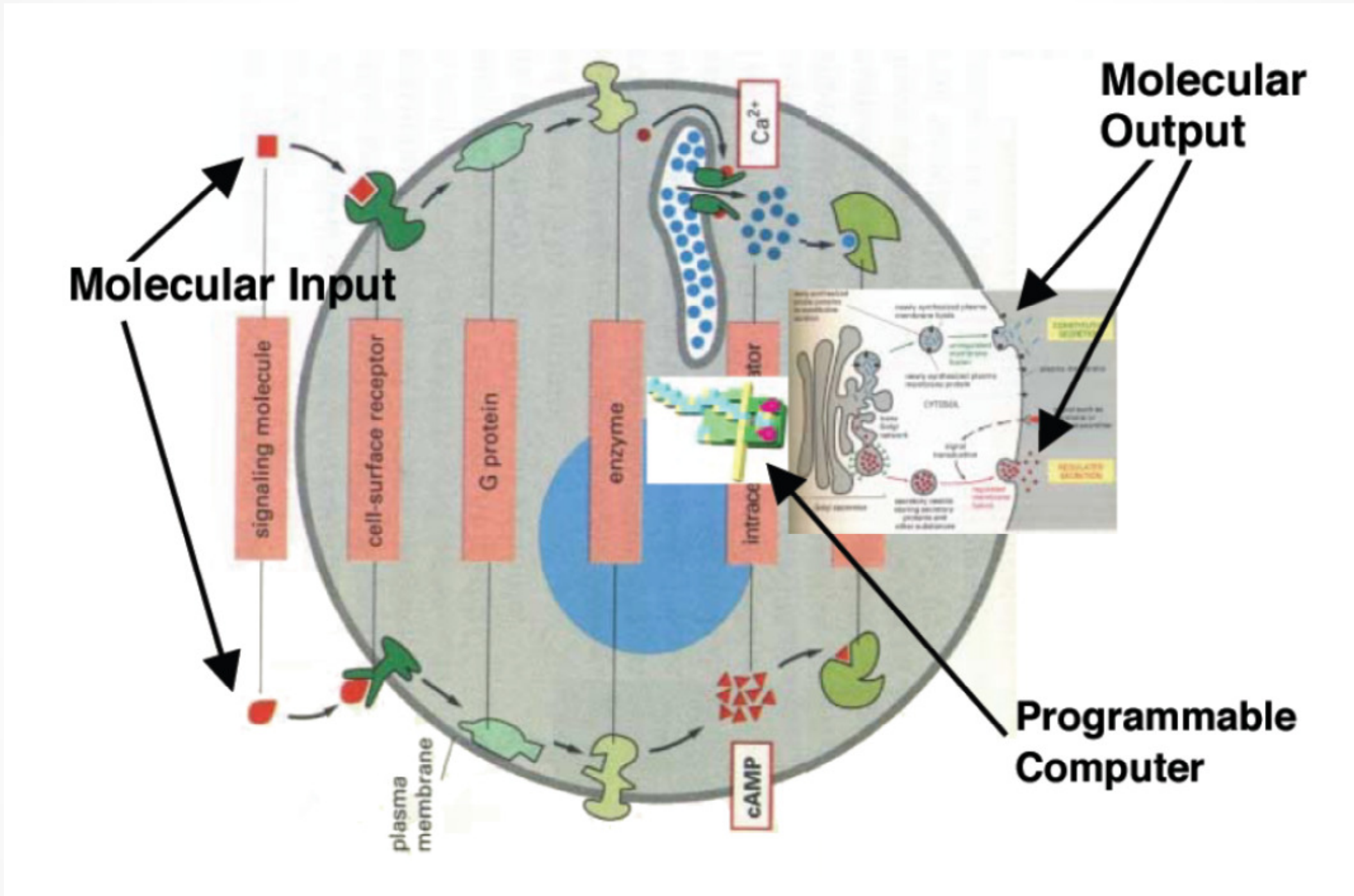


Fig. 1 Medicine in 2050: "Doctor in a Cell"

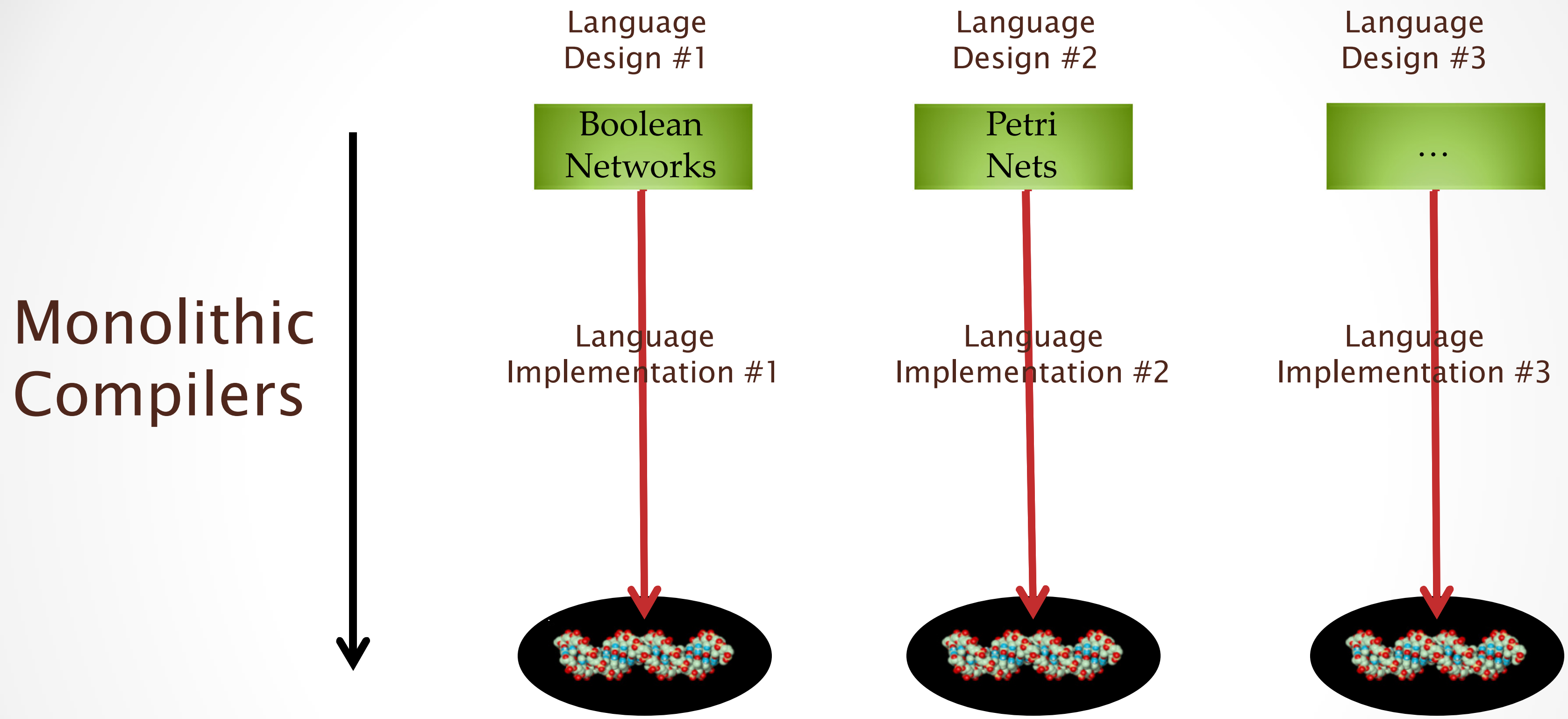
Ehud Shapiro

Rivka Adar
Kobi Benenson
Gregory Linshitz
Aviv Regev
William Silverman

Molecules and computation

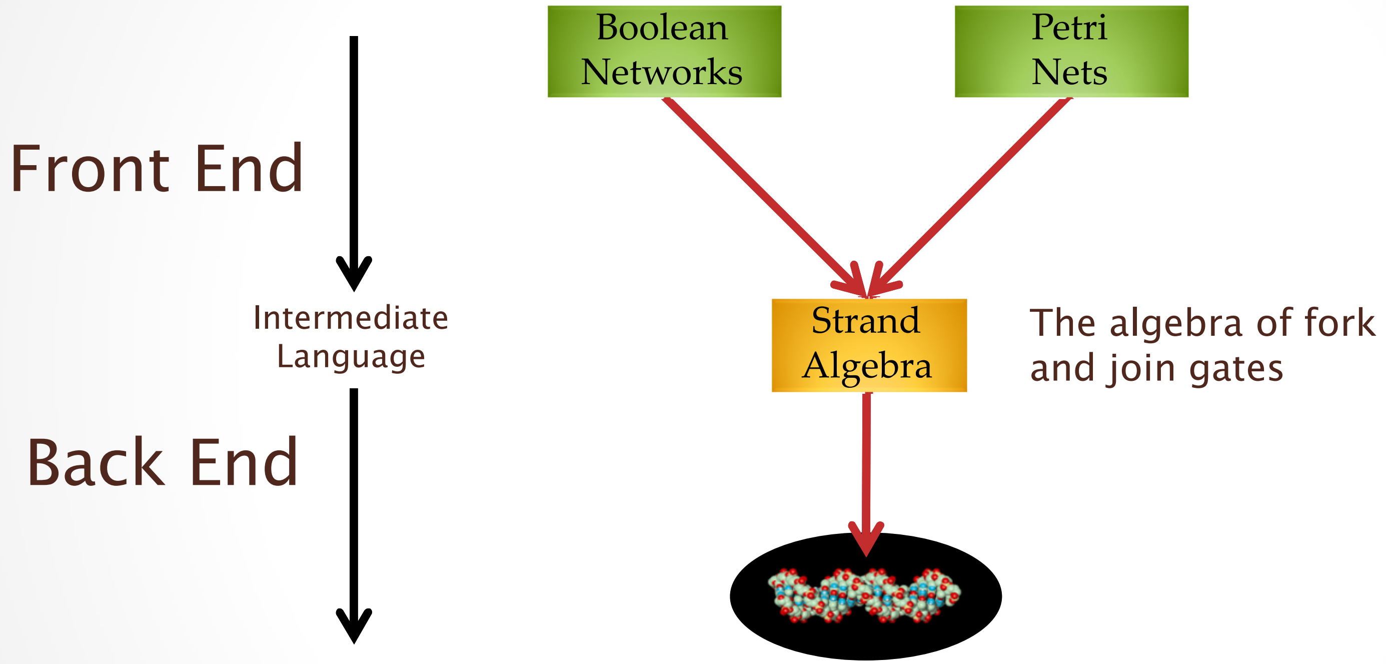
DNA Compilation

Compilers



Monolithic
Compilers
Language
Design #1
Boolean
Networks
Language
Implementation #1
Language
Design #2
Petri
Nets
Language
Implementation #2
Language
Design #3
...
Language
Implementation #3

Intermediate Languages

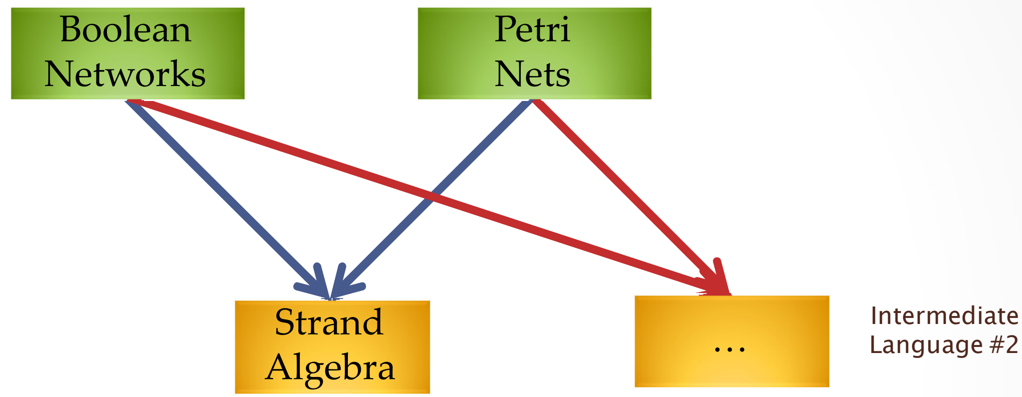


Front Ends

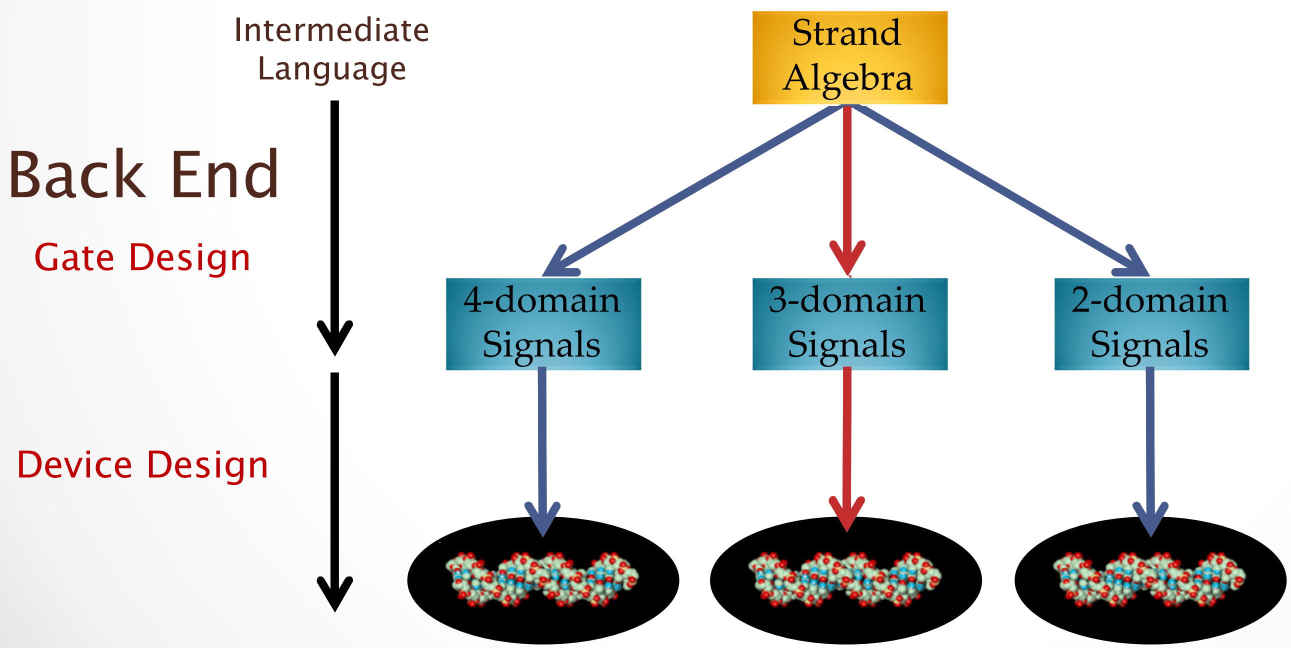
Front End

Circuit Design

Intermediate Language



Back Ends


Intermediate
Language
Back End
Gate Design
Device Design
Strand
Algebra
4-domain
Signals
3-domain
Signals
2-domain
Signals

The Actual Tutorial

(The Back End Work)

Gates to Structures

- Visual DSD (Andrew Phillips)
  - A tool for exploring, e.g., gate designs

- Fork gate: the program

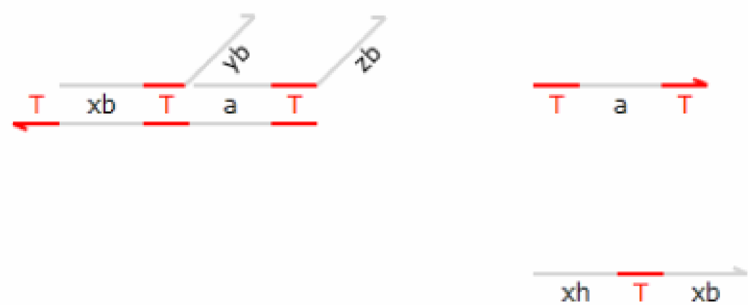
```
directive sample 5000.0 1000
directive plot sum(<_ T^ xb>); sum(<_ T^ yb>); sum(<_ T^ zb>)
def scaling = 1000
def bind = 0.0003/(float_of_int scaling) (* /nM/s *) (* =3*10^5 /M/s *)
def unbind =  0.1126 (* /s *)
new T@bind,unbind

def F1x2(N,Xb,Yb,Zb) =
new a
( N * T^:[Xb T^]<Yb>:[a T^]<Zb>
| N * <T^ a T^>
)

(   F1x2(10*scaling,xb,yb,zb)
|  (1*scaling)* <xh T^ xb>
)
```

Gates to Structures

- Visual DSD (Andrew Phillips)
  - A tool for exploring, e.g., gate designs
- Fork gate: the structures



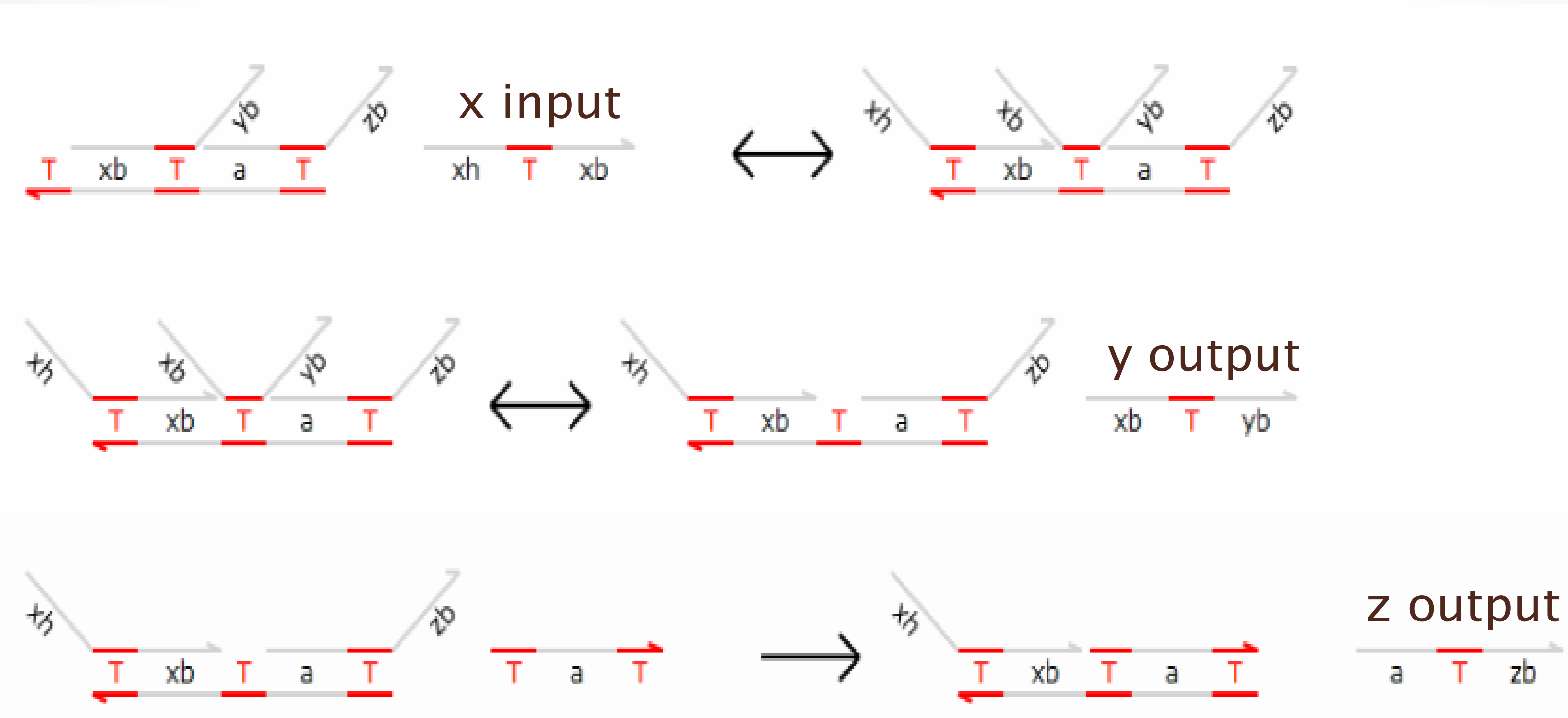
$x \rightarrow y + z$ gate

x input

Gates to Structures

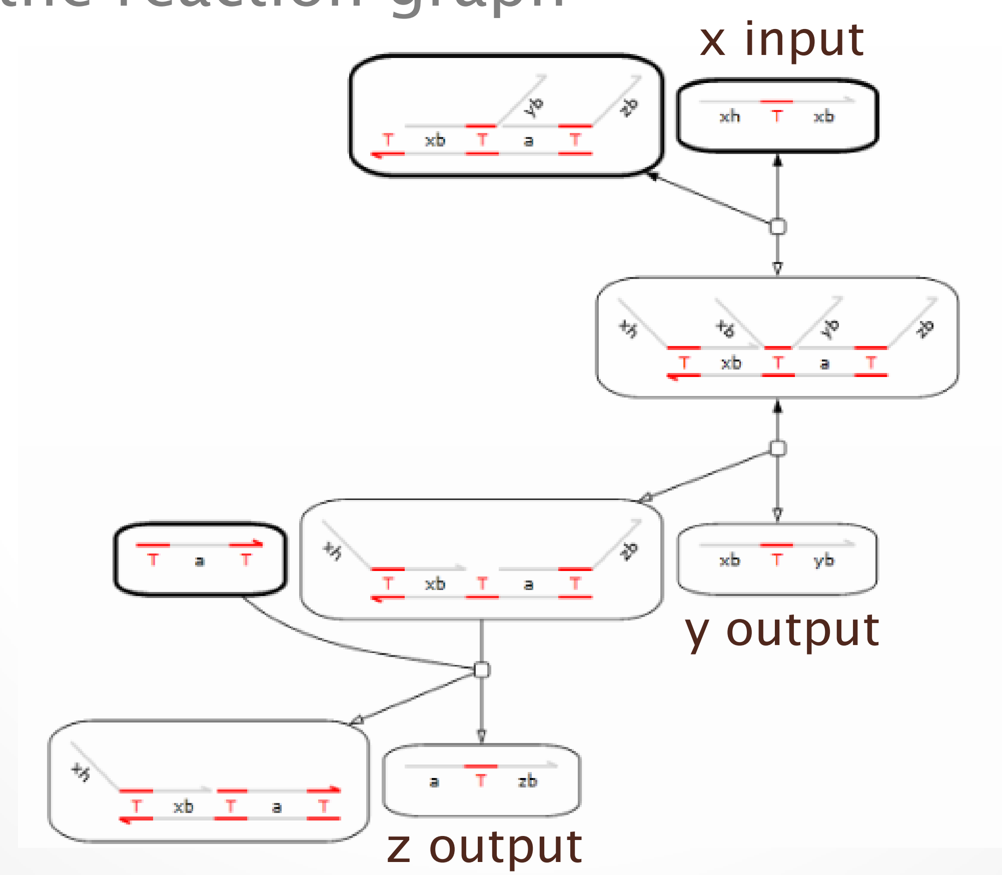
- Visual DSD (Andrew Phillips)
  - A tool for exploring, e.g., gate designs

- Fork gate: the reactions



Gates to Structures

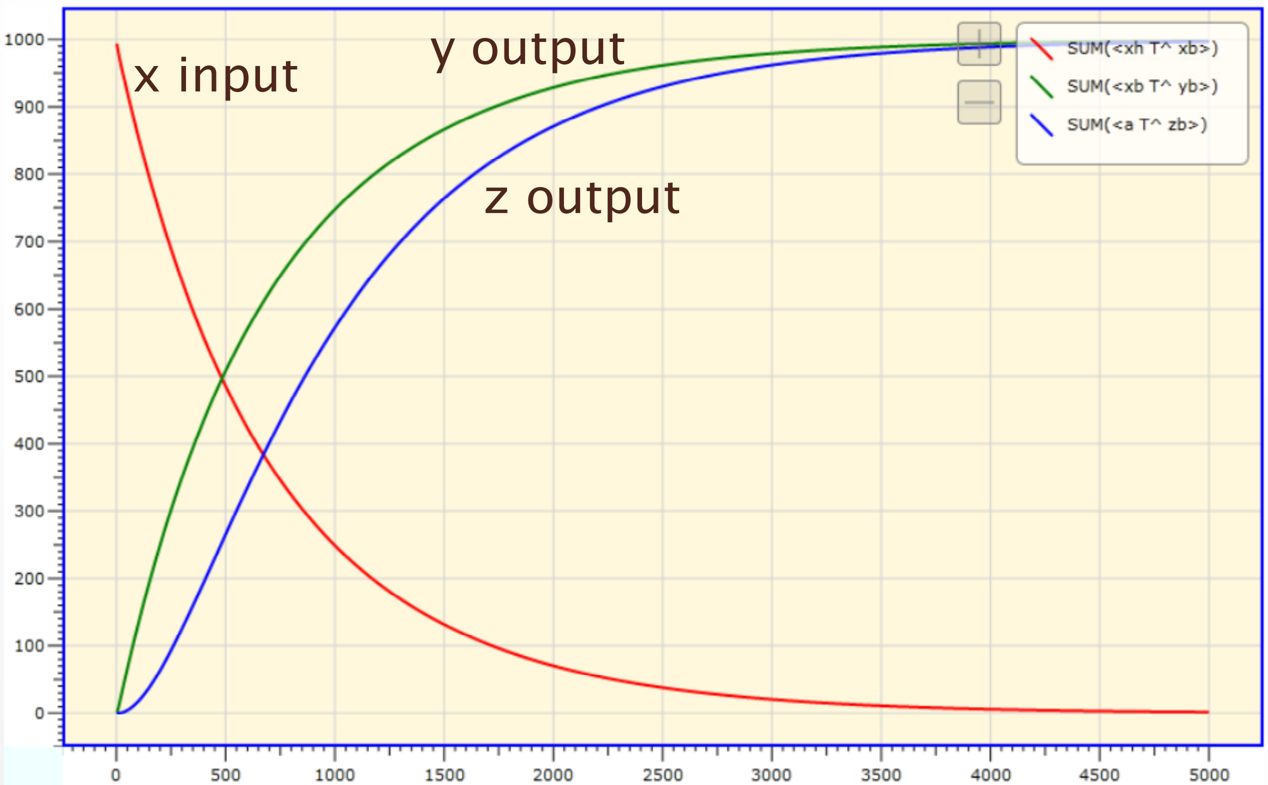
- Visual DSD (Andrew Phillips)
  - A tool for exploring, e.g., gate designs
- Fork gate: the reaction graph



Gates to Structures

- Visual DSD (Andrew Phillips)
  - A tool for exploring, e.g., gate designs

- Fork gate: the behavior

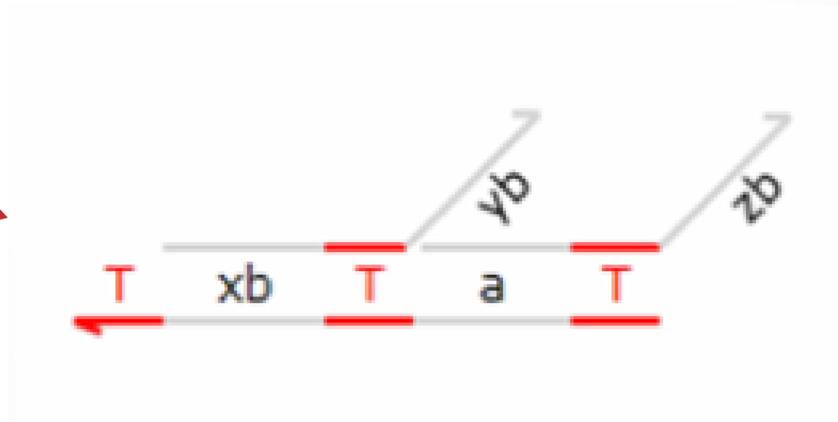


Gates to Structures

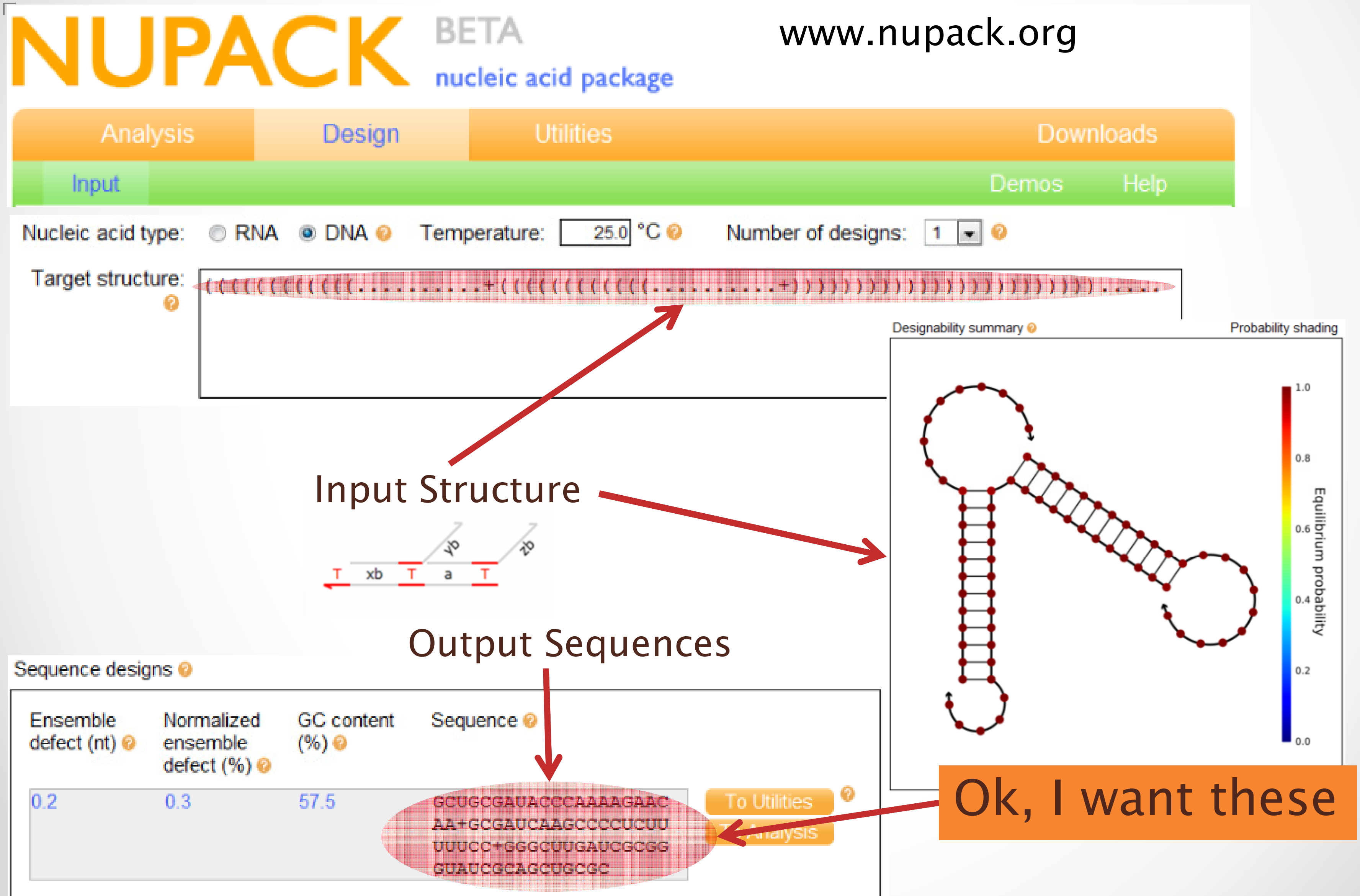
- Visual DSD (Andrew Phillips)
  - A tool for exploring, e.g., gate designs

- Fork gate: check

Ok, I want this

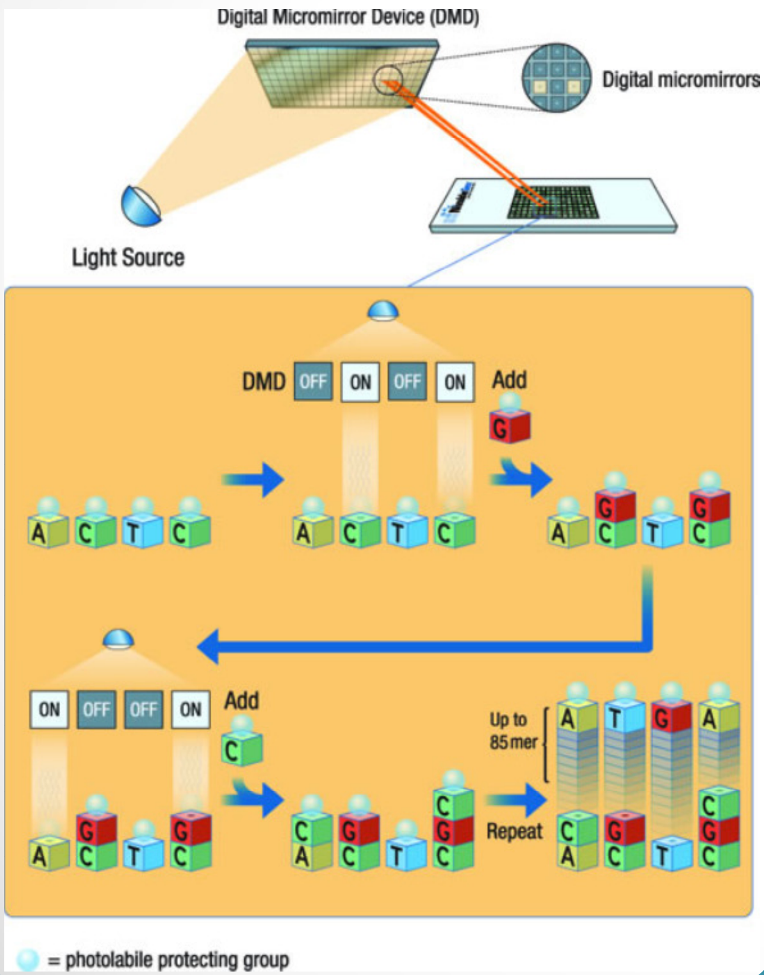


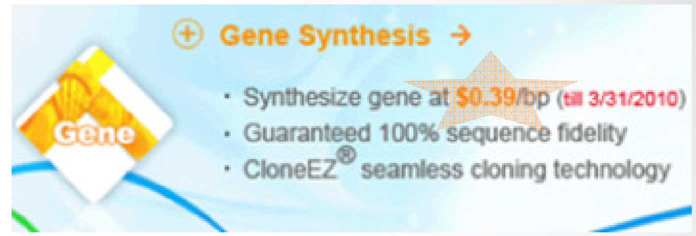
Structures to Sequences



Sequences to DNA

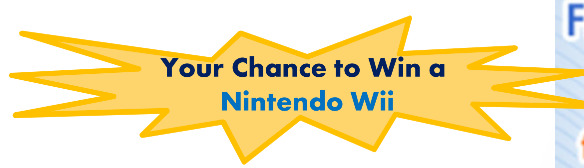
Oligonucleotide Synthesis





Fastest tunaround time for less money!

Standard gene synthesis from 0.36 €/bp in just 8 days



SameDay® Oligo Service

Only £0.57 GBP / Base!

Base Pricing

Synthesis Scale	Price	
25 nmole DNA Oligo	£0.25 GBP / Base	Order
100 nmole DNA oligo	£0.45 GBP / Base	Order
250 nmole DNA oligo	£0.80 GBP / Base	Order
1 µmole DNA oligo	£1.60 GBP / Base	Order
5 µmole DNA oligo	£7.50 GBP / Base	Order
10 µmole DNA oligo	£14.50 GBP / Base	Order

Custom DNA/RNA
Pricing (USD)

DNA(mg)	Desalted	Purified
15	$700	$1,050
50	$1,200	$1,450
100	$1,500	$1,800
250	$2,000	$2,400
500	$2,900	$3,400
1000	$4,550	$5,400
5000	$9,000	$10,700

RNA(mg)	Desalted	Purified
5	$1,500	$1,925
15	$1,950	$2,490
50	$2.050	$2,625
100	$2,575	$3,575
250	$4,575	$5,725
500	$7,900	$9,190
1000	$13,900	$15,900
5000		$37,125

Please inquire for larger quantities

Experiments

How are they Actually Done?

Web Order

#	Name	Product	Usually Ships In	Length	Purification	Guaranteed Yield	Sequence	Price
#1	Gb_bot	250 nmole DNA oligo	1 business day	69	Standard Desalting	15 ODs = 20.9 nmoles = [illegible] µgrams	5'- GAG GAG GTT GTG AAG TAA TGC GTG AGA TGT GAT TGT GTT ATG GTG AGG GTA AA A GGT GAA TTGGAG GAG -3'	$37.95
#2	B tx	250 nmole DNA oligo	1 business day	21	Standard Desalting	15 ODs = 81.1 nmoles = 505.9 µgrams	5'- CAA TTC ACC TTT TAC CCT CAC -3'	$11.55
#3	x tb	250 nmole DNA oligo	1 business day	21	Standard Desalting	15 ODs = 73.3 nmoles = 460.9 µgrams	5'- CAT AAC ACA ATC ACA TCT CAC -3'	$11.55
#4	b tb B	250 nmole DNA oligo	1 business day	36	Standard Desalting	15 ODs = 46.5 nmoles = 501.8 µgrams	5'- GCA TTA CTT CAC AAC CTC CTC CAA TTC ACC TTT TAC -3'	$19.80
#5	B	250 nmole DNA oligo	1 business day	15	Standard Desalting	10 ODs = 73.7 nmoles = 329.1 µgrams	5'- CAA TTC ACC TTT TAC -3'	$8.25
#6	GB_bot	250 nmole DNA oligo	1 business day	21	Standard Desalting	15 ODs = 67.3 nmoles = 444.9 µgrams	5'- GTG AGG GTA AAA GGT GAA TTG-3'	$11.55
#7	Gt	250 nmole DNA oligo	1 business day	27	Standard Desalting	15 ODs = 63.2 nmoles = 509.1 µgrams	5'- TCT CAC GCA TTA CTT CAC AAC CTC CTC -3'	$14.85
#8	tx x	250 nmole DNA oligo	1 business day	21	Standard Desalting	15 ODs = 73.5 nmoles = 461 µgrams	5'- CCT CAC CAT AAC ACA ATC ACA -3'	$11.55

SubTotal	$127.05 USD
Shipping and Handling	$16.00 USD
Tax	$12.39 USD
Total	$155.44 USD

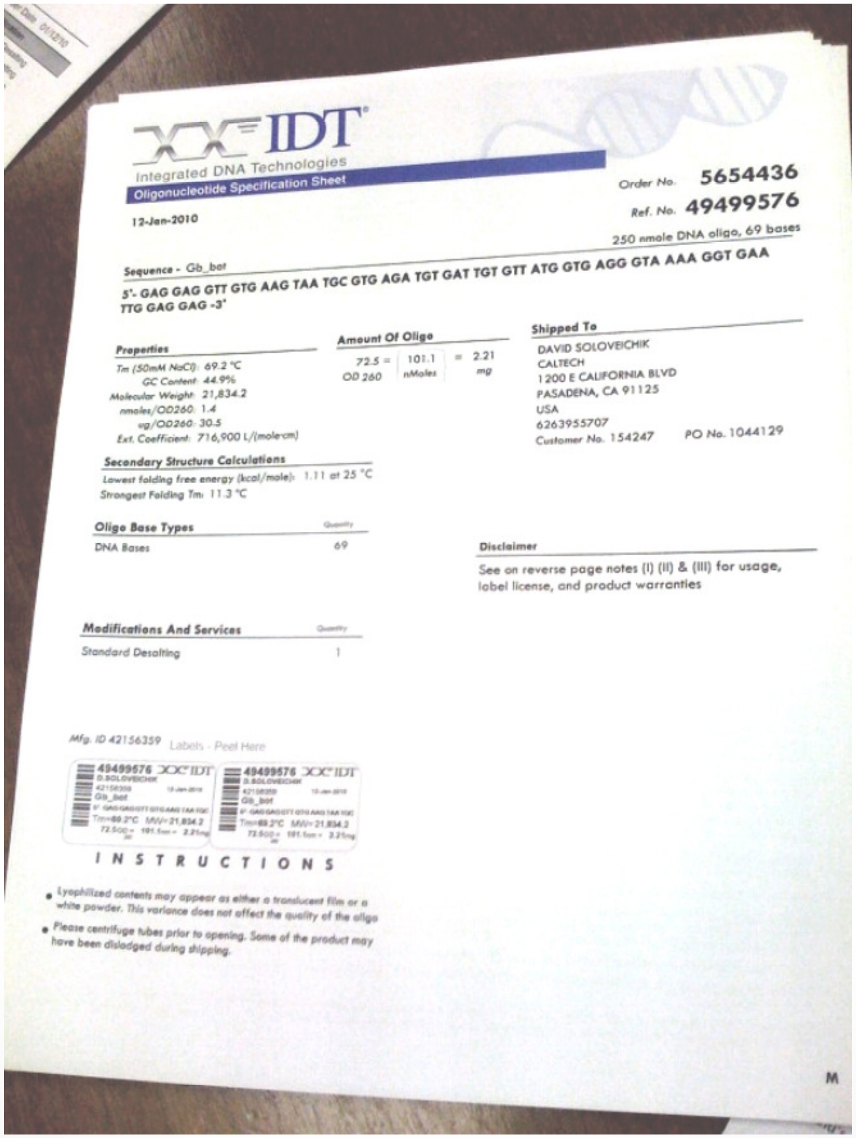
Wait 24 Hours

• • •

DNA by Mail


IDT
INTEGRATED DNA TECHNOLOGIES, INC.
1710 Commercial Park • Coralville, IA 52241-2760 • USA
6868 Nancy Ridge Drive • San Diego, CA 92121-2232 • USA
Interleuvenlaan 12A • 3001 Leuven • Belgium
800-328-2661 (US only)
Custom
Oligonucleotide
Synthesis
Innovation and
Precision in
Nucleic Acid
Synthesis

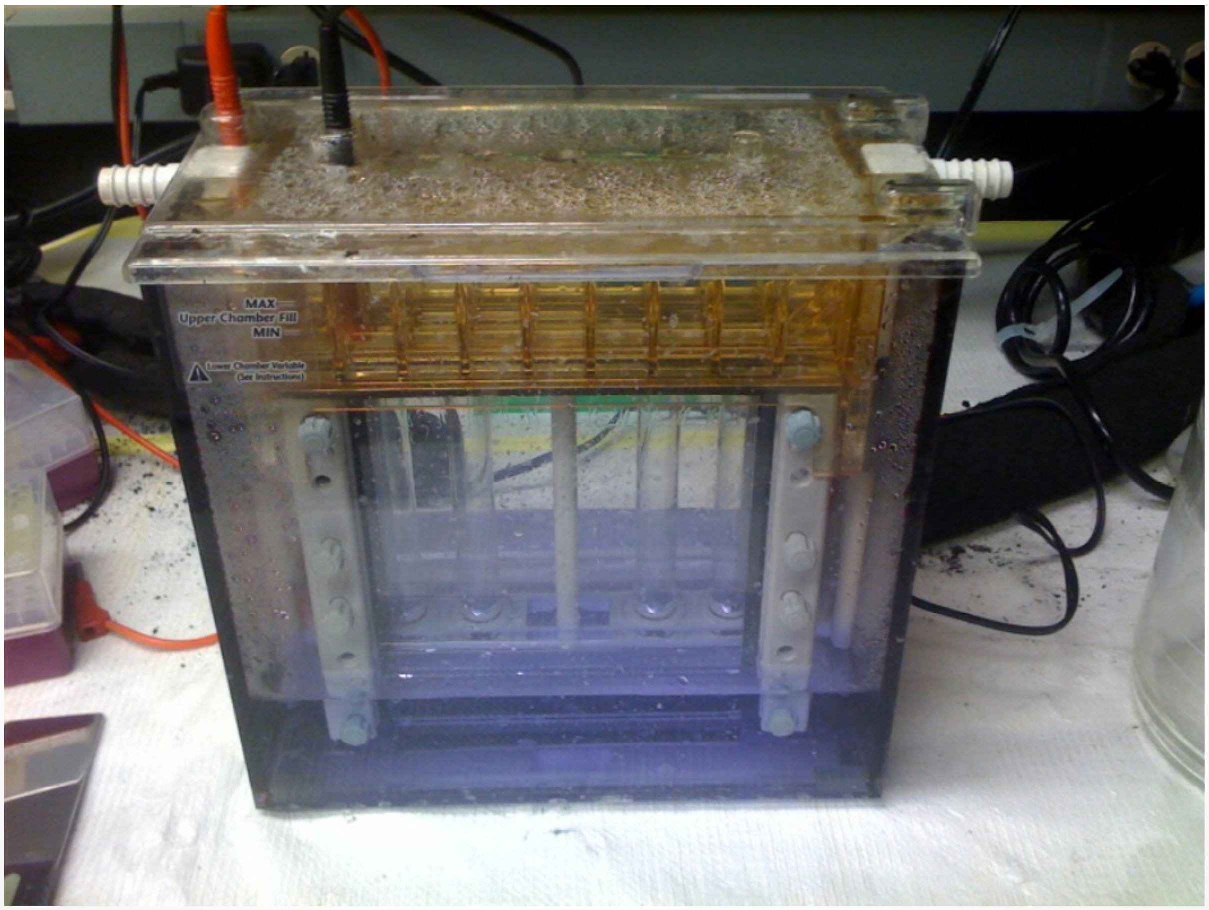
Spec Sheet


IDT
Integrated DNA Technologies
Oligonucleotide Specification Sheet
12-Jan-2010
Order No. 5654436
Ref. No. 49499576
250 nmole DNA oligo, 69 bases
Sequence - Gb_bot
5'- GAG GAG GTT GTG AAG TAA TGC GTG AGA TGT GAT TGT GTT ATG GTG AGG GTA AAA GGT GAA TTG GAG GAG -3'
Properties
Tm (50mM NaCl): 69.2 °C
GC Content: 44.9%
Molecular Weight: 21,834.2
nmoles/OD260: 1.4
ug/OD260: 30.5
Ext. Coefficient: 716,900 L/(mole·cm)
Amount Of Oligo
72.5 OD 260 = 101.1 nMoles = 2.21 mg
Shipped To
DAVID SOLOVEICHIK
CALTECH
1200 E CALIFORNIA BLVD
PASADENA, CA 91125
USA
6263955707
Customer No. 154247
PO No. 1044129
Secondary Structure Calculations
Lowest folding free energy (kcal/mole): 1.11 at 25 °C
Strongest Folding Tm: 11.3 °C
Oligo Base Types
DNA Bases 69
Disclaimer
See on reverse page notes (I) (II) & (III) for usage, label license, and product warranties
Modifications And Services
Standard Desalting 1
Mfg. ID 42156359
Labels - Peel Here
INSTRUCTIONS
Lyophilized contents may appear as either a translucent film or a white powder. This variance does not affect the quality of the oligo
Please centrifuge tubes prior to opening. Some of the product may have been dislodged during shipping.
M

Add Water

Put DNA into Gel

- Polyacrylamide gel electrophoresis (PAGE)
- Sorts DNA strands by length

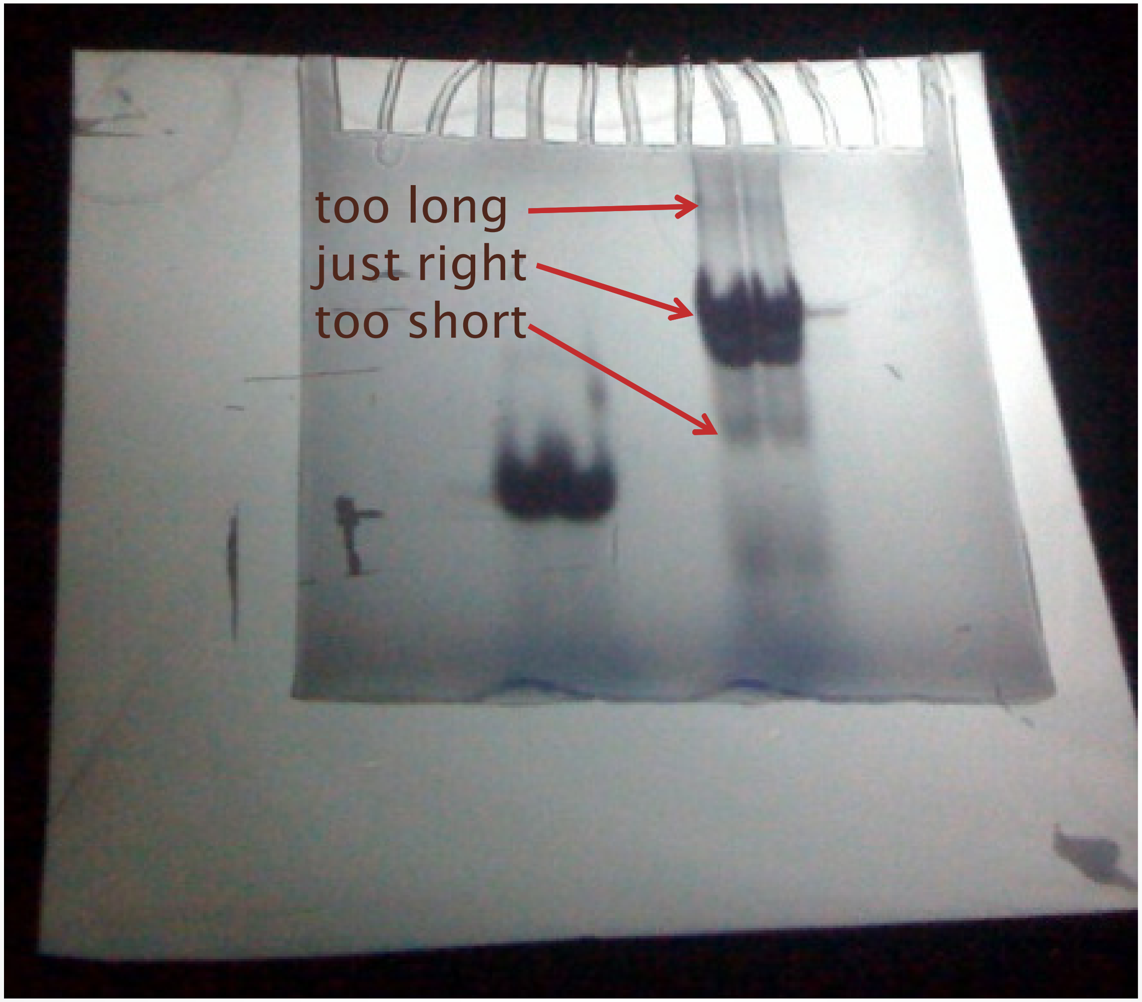


Wait 6 Hours

• • •

Get DNA out of Gel

- Find DNA with ultraviolet light. Cut it out.



Wait 12 Hours

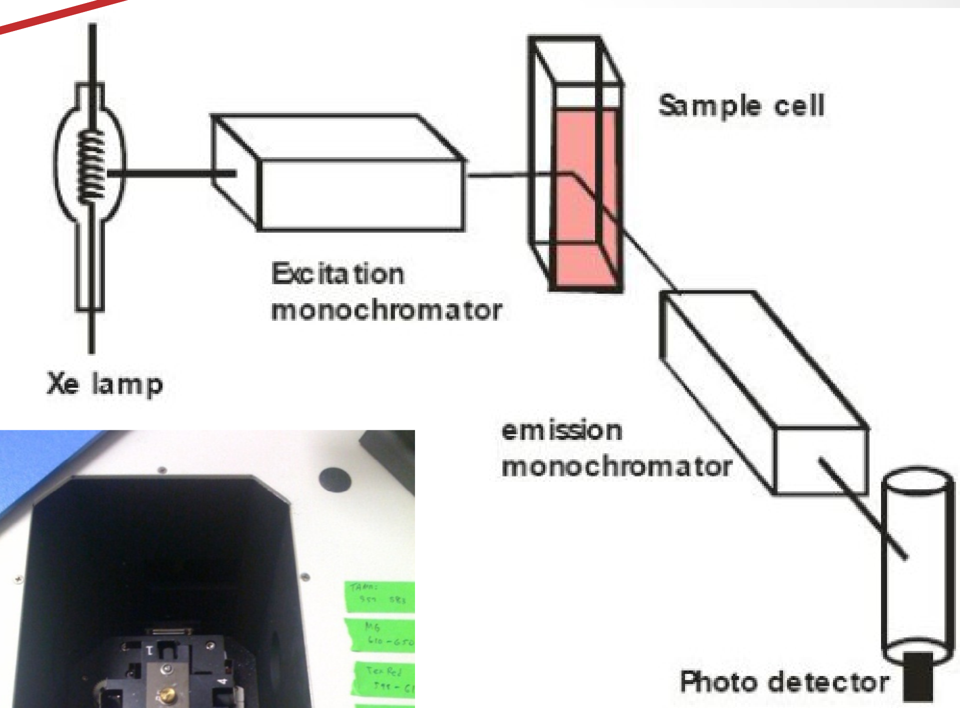
• • •

Mix DNA Up

- Screaming for robotic automation

Spectrofluorometer

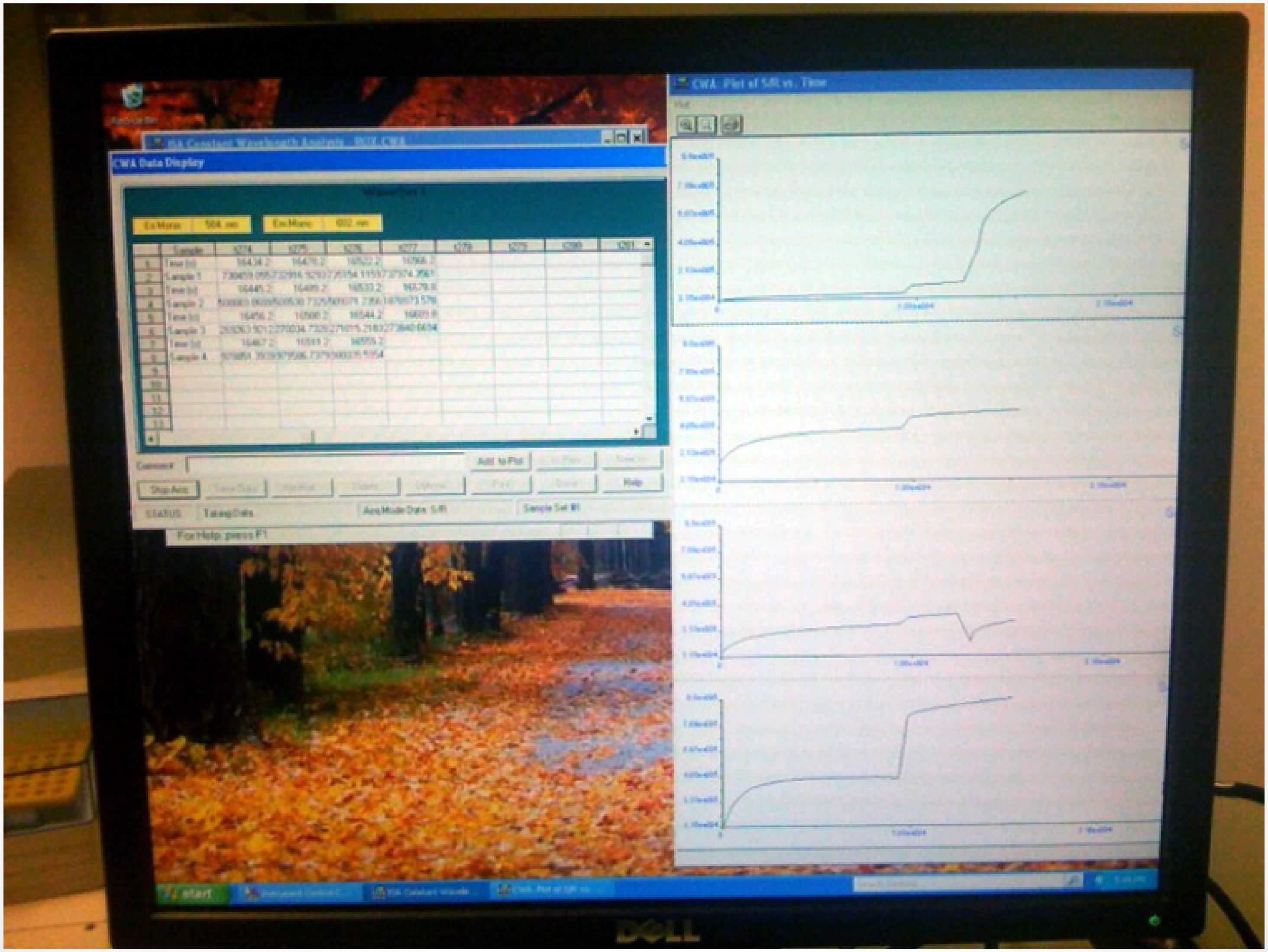
- Fluorescence is your 'print' statement



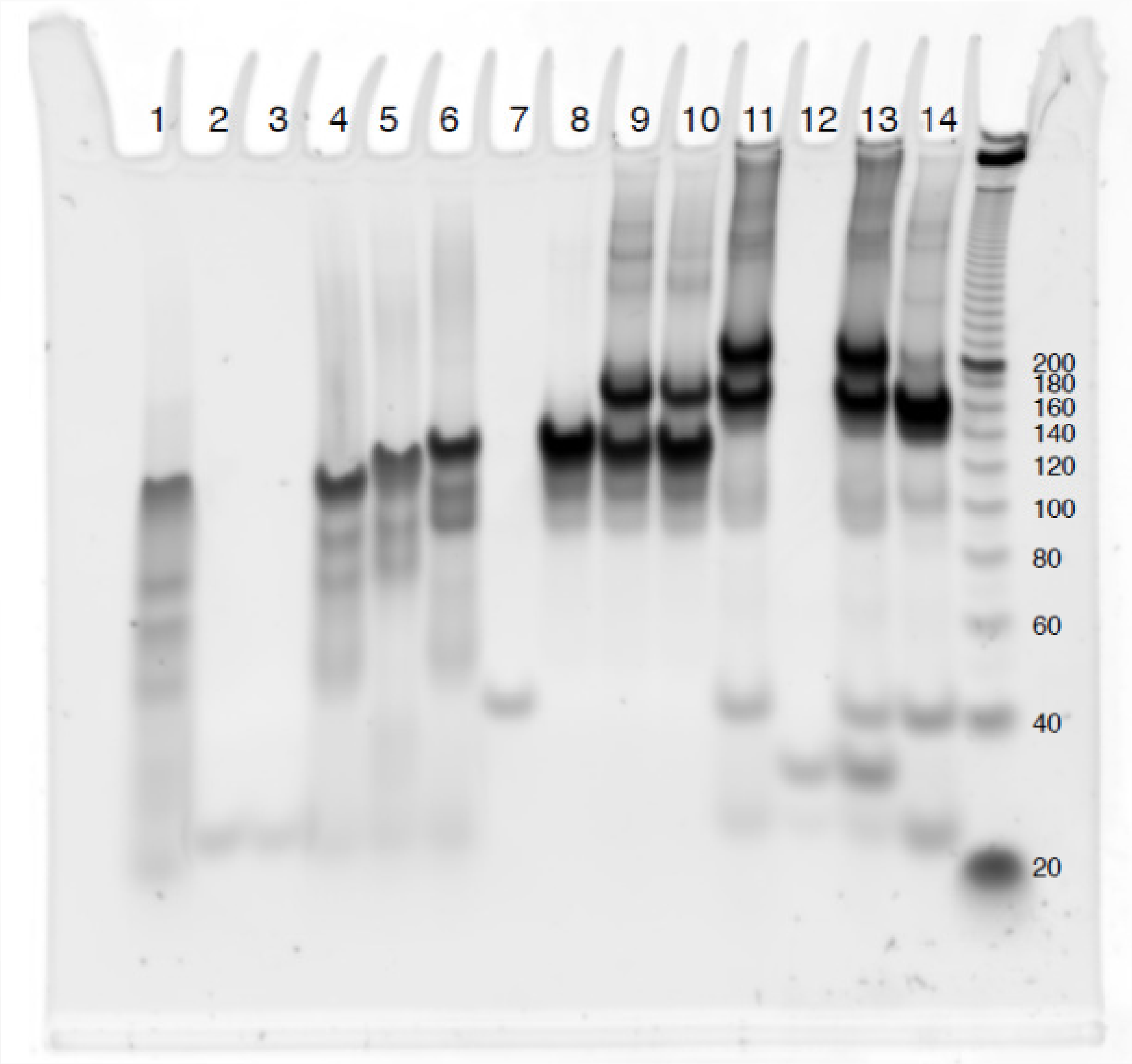
Go To Lunch

• • •

Execution Trace



Core Dump



Repeat

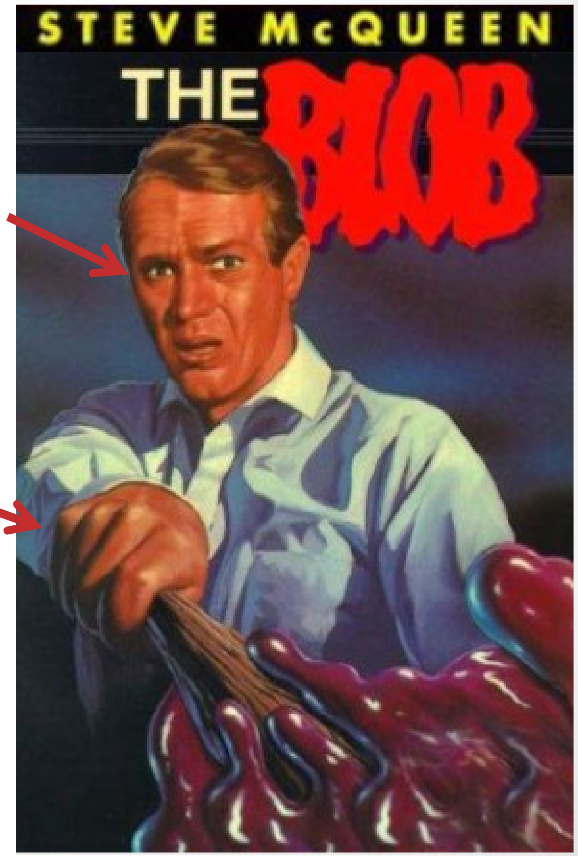
• • •

Health and Safety

- Don’t try this at home
  - (Although you could)
- Latex gloves, UV glasses
  - Fear the Gel (acrylamide)
  - Fear the Light (UV)
- Otherwise safe
  - No smells
  - No fires
  - No biohazards
  - No life forms
- Most complex machines:
  - Gel machine
  - Fluorometer
  - Atomic force microscope
- Most dangerous activity:
  - Replacing the light bulb in the fluorometer (hot; may explode)

Not wearing glasses

Not wearing gloves



Acknowledgments

- Illustrations
  - John Reif, Duke
  - Ned Seeman, NYU
  - Erik Winfree, Caltech
  - Bernard Yurke, Boise State
  - Wikipedia
  - YouTube
- David Soloveichik